(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 October 2003 (30.10.2003)

PCT

(10) International Publication Number
WO 03/088910 A2

(51) International Patent Classification[7]: A61K

(21) International Application Number: PCT/US03/11867

(22) International Filing Date: 15 April 2003 (15.04.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

10/123,568	15 April 2002 (15.04.2002)	US
10/123,731	15 April 2002 (15.04.2002)	US
60/373,366	16 April 2002 (16.04.2002)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 10/123,568 (CIP)
Filed on 15 April 2002 (15.04.2002)

(71) Applicant *(for all designated States except US)*: **RIGEL PHARMACEUTICALS, INC.** [US/US]; 1180 Veterans Blvd., South San Francisco, CA 94080 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **HITOSHI, Yasumichi** [JP/US]; 331 Calfippe Court, Brisbane, CA 94005 (US). **JENKINS, Yonchu** [US/US]; 6529 Ascot Drive, Oakland, CA 94611 (US).

(74) Agents: **KELLY, Beth, L.** et al.; TOWNSEND AND TOWNSEND AND CREW LLP, Two Embarcadero Center, 8th Floor, San Francisco, CA 94111-3834 (US).

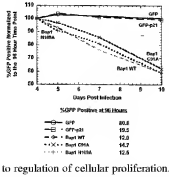
(81) Designated States *(national)*: AE, AG, AL, AM, AT (utility model), AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ (utility model), CZ, DE (utility model), DE, DK (utility model), DK, DM, DZ, EC, EE (utility model), EE, ES, FI (utility model), FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK (utility model), SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO,

[Continued on next page]

(54) Title: METHODS OF ASSAYING FOR CELL CYCLE MODULATORS

Expression of Bap1 WT and Protease Mutants is Antiproliferative In HeLa Cells



(57) Abstract: The present invention relates to regulation of cellular proliferation. More particularly, the present invention is directed to nucleic acids encoding BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DEAD/H box polypeptide 21 (DDX21), serine threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, or ERCC1, which are involved in modulation of cell cycle arrest. The invention further relates to methods for identifying and using agents, including small molecule chemical compositions, antibodies, peptides, cyclic peptides, nucleic acids, RNAi, antisense nucleic acids, and ribozymes, that modulate cell cycle arrest via modulation of BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DEAD/H box polypeptide 21 (DDX21), serine threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, or ERCC1, as well as to the use of expression profiles and compositions in diagnosis and therapy related to cell cycle regulation and modulation of cellular proliferation, e.g., for treatment of cancer and other diseases of cellular proliferation.

WO 03/088910 A2

SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations of inventorship (Rule 4.17(iv)) for US only*

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHODS OF ASSAYING FOR CELL CYCLE MODULATORS

CROSS-REFERENCES TO RELATED APPLICATIONS

This application claims the benefit of priority of each of the following: U.S. application serial number 10/123,568 filed April 15, 2002; U.S. application serial number 10/123,731 filed April 15, 2002; and U.S. provisional application serial number 60/373,366 filed April 16, 2002. Each of the following applications are herein incorporated by reference for all purposes: U.S. application serial number 10/123,568 filed April 15, 2002; U.S. application serial number 10/123,731 filed April 15, 2002; and U.S. provisional application serial number 60/373,366 filed April 16, 2002.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

Not applicable.

FIELD OF THE INVENTION

The present invention relates to regulation of cellular proliferation. More particularly, the present invention is directed to nucleic acids encoding BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase (G6PD), HCDR-3, DEAD/H box polypeptide 21 (DDX21), serine threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, or ERCC1, which are involved in modulation of cell cycle arrest. The invention further relates to methods for identifying and using agents, including small molecule chemical compositions, antibodies, peptides, cyclic peptides, nucleic acids, RNAi, antisense nucleic acids, and ribozymes, that modulate cell cycle arrest via modulation of BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DEAD/H box polypeptide 21

(DDX21), serine threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, or ERCC1, as well as to the use of expression profiles and compositions in diagnosis and therapy related to cell cycle regulation and modulation of cellular proliferation, e.g., for treatment of cancer and other diseases of cellular proliferation.

BACKGROUND OF THE INVENTION

Cell cycle regulation plays a critical role in neoplastic disease, as well as disease caused by non-cancerous, pathologically proliferating cells. Normal cell proliferation is tightly regulated by the activation and deactivation of a series of proteins that constitute the cell cycle machinery. The expression and activity of components of the cell cycle can be altered during the development of a variety of human disease such as cancer, cardiovascular disease, psoriasis, where aberrant proliferation contributes to the pathology of the illness. There are genetic screens to isolate important components for cell cycle regulation using different organisms such as yeast, worms, flies, etc. However, involvement of a protein in cell cycle regulation in a model system is not always indicative of its role in cancer and other proliferative disease. Thus, there is a need to establish screening for understanding human diseases caused by disruption of cell cycle regulation. Identifying proteins, their ligands and substrates, and downstream signal transduction pathways involved in cell cycle regulation and neoplasia in humans is important for developing therapeutic regents to treat cancer and other proliferative diseases.

BRIEF SUMMARY OF THE INVENTION

The present invention therefore provides nucleic acids encoding BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DEAD/H box polypeptide 21 (DDX21), serine threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, or ERCC1, which are involved in modulation of cell cycle arrest in tumor cells and other pathologically proliferating cells. The invention therefore provides methods of screening for compounds, e.g., small organic molecules, antibodies, peptides, cyclic peptides, nucleic acids, antisense molecules, RNAi, and ribozymes, that are capable of modulating cellular proliferation and/or cell cycle regulation, e.g., either inhibiting cellular proliferation, or activating apoptosis. Therapeutic and diagnostic methods and reagents are also provided.

Modulators of BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DEAD/H box polypeptide 21 (DDX21), serine threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, or ERCC1 are therefore useful in treatment of cancer and other proliferative diseases.

One embodiment of the present invention provides a method for identifying a compound that modulates cell cycle arrest. A cell comprising an BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DEAD/H box polypeptide 21 (DDX21), serine threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, or ERCC1 polypeptide or fragment thereof is contacted with the compound. The BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DEAD/H box polypeptide 21 (DDX21), serine threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, or ERCC1 polypeptide or fragment thereof may be encoded by a nucleic acid that hybridizes under stringent conditions to a nucleic acid encoding a polypeptide having an amino acid sequence of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, or 28. The chemical or phenotypic effect of the compound upon the cell comprising the BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DEAD/H box polypeptide 21 (DDX21), serine threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, or ERCC1 polypeptide or fragment thereof is determined, thereby identifying a compound that modulates cell cycle arrest. The chemical or phenotypic effect may be determined by measuring enzymatic activity of the BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1

(UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DEAD/H box polypeptide 21 (DDX21), serine threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, or ERCC1 polypeptide. The chemical or phenotypic effect may be determined by measuring cell cycle arrest. The cell cycle arrest may be measured by assaying DNA synthesis or fluorescent marker level. DNA synthesis may be measured by ^{3}H thymidine incorporation, BrdU incorporation, or Hoescht staining. The fluorescent marker may be a cell tracker dye or green fluorescent protein. Modulation may be activation of cell cycle arrest or activation of cancer cell cycle arrest. The host cell may be a cancer cell. The cancer cell may be a breast, prostate, colon, or lung cancer cell. The cancer cell may be a transformed cell line, such as, for example, PC3, H1299, MDA-MB-231, MCF7, A549, or HeLa. The cancer cell may be p53 null, p53 mutant, or p53 wild-type. The polypeptide may recombinant. The polypeptide may be encoded by a nucleic acid comprising a sequence of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, or 27. The compound may be an antibody, an antisense molecule, a small organic molecule, a peptide, or a circular peptide.

Another embodiment of the invention provides a method for identifying a compound that modulates cell cycle arrest. The compound is contacted with an BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DEAD/H box polypeptide 21 (DDX21), serine threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, or ERCC1 polypeptide or a fragment thereof, the BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DEAD/H box polypeptide 21 (DDX21), serine threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, or ERCC1 polypeptide or fragment thereof. The BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DEAD/H box polypeptide 21 (DDX21), serine threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, or ERCC1 polypeptide or a fragment

thereof may be encoded by a nucleic acid that hybridizes under stringent conditions to a nucleic acid encoded by a polypeptide comprising an amino acid sequence of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, or 28. The physical effect of the compound upon the BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DEAD/H box polypeptide 21 (DDX21), serine threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, or ERCC1 polypeptide is determined. The chemical or phenotypic effect of the compound upon a cell comprising an BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DEAD/H box polypeptide 21 (DDX21), serine threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, or ERCC1 polypeptide or fragment thereof is determined, thereby identifying a compound that modulates cell cycle arrest.

Yet another embodiment of the invention provides a method of modulating cell cycle arrest in a subject. A therapeutically effective amount of a compound identified according to one of the methods described above is administered to the subject. The subject may be a human. The subject may have cancer. The compound may inhibit cancer cell proliferation.

Even another embodiment of the invention provides a method of modulating cell cycle arrests in a subject. A therapeutically effective amount of a BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DEAD/H box polypeptide 21 (DDX21), serine threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, or ERCC polypeptide is administered to the subject. The BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DEAD/H box polypeptide 21 (DDX21), serine

threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, or ERCC1 polypeptide may be encoded by a nucleic acid that hybridizes under stringent conditions to a nucleic acid encoding a polypeptide having an amino acid sequence of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, or 28.

A further embodiment of the invention provides a method of modulating cell cycle arrest in a subject. A therapeutically effective amount of a BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DEAD/H box polypeptide 21 (DDX21), serine threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, or ERCC1 polypeptide is administered to the subject. The BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DEAD/H box polypeptide 21 (DDX21), serine threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, or ERCC1 polypeptide may be encoded by a nucleic acid that hybridizes under stringent conditions to a nucleic acid encoding a polypeptide having an amino acid sequence of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, or 28.

Other embodiments and advantages of the present invention will be apparent from the detailed description that follows.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 provides a nucleotide (SEQ ID NO:1) and amino acid sequence (SEQ ID NO:2) of human BAP-1.

Figure 2 provides an illustration of the relevant domains of BAP-1, including the ubiquitin hydrolase domain and the DNA binding domain. Also shown is the BAP-1 functional hit (G3-2D8) isolated in the retroviral screen. The functional hit is in the antisense orientation.

Figure 3 illustrates cell tracker assay data demonstrating that GFP-fused BAP-1 is antiproliferative in A549 cells. The BAP-1 construct is the functional hit isolated in the retroviral screen. Figure 3 top left illustrates fluorescence analysis of green fluorescent protein (GFP) infected A549.tTA control cells. Figure 3 top right illustrates cell tracker assay

data from GFP infected A549.tTA control cells. Figure 3 lower left illustrates fluorescence analysis of BAP-1 infected A549.tTA cells. Figure 3 lower right illustrates cell tracker assay date from BAP-1 infected A549.tTA cells.

Figure 4 provides a nucleotide (SEQ ID NO:3) and amino acid sequence (SEQ ID NO:4) of human NP95.

Figure 5 provides an illustration of the relevant domains of NP95, including the ubiquitin like domain, the zinc finger domain, the nuclear protein domain, and the ubiquitin ligase domain.

Figure 6 illustrates cell tracker assay data demonstrating that GFP-fused NP95 is antiproliferative in A549. The NP-95 construct is the functional hit isolated in the retroviral screen. Figure 6 top left illustrates fluorescence analysis of green fluorescent protein (GFP) infected A549.tTA control cells. Figure 6 top right illustrates cell tracker assay data from GFP infected A549.tTA control cells. Figure 6 lower left illustrates fluorescence analysis of NP95 infected A549.tTA cells. Figure 6 lower right illustrates cell tracker assay date from NP95 infected A549.tTA cells.

Figure 7 provides a nucleotide (SEQ ID NO:5) and amino acid (SEQ ID NO:6) sequence of human FANCA.

Figure 8 provides a nucleotide (SEQ ID NO:7) and an amino acid (SEQ ID NO:8) sequence of human DDX9.

Figure 9 provides a nucleotide (SEQ ID NO:9) and an amino acid (SEQ ID NO:10) sequence of human IGF1R.

Figure 10 provides a nucleotide (SEQ ID NO:11) and an amino acid (SEQ ID NO:12) sequence of human UBE2V1.

Figure 11 provides a nucleotide (SEQ ID NO:13) and an amino acid (SEQ ID NO:14) sequence of human aldehyde dehydrogenase.

Figure 12 provides a nucleotide (SEQ ID NO:15) and an amino acid (SEQ ID NO:16) sequence of human pyruvate kinase.

Figure 13 provides a nucleotide (SEQ ID NO:17) and an amino acid (SEQ ID NO:18) sequence of human G6PD.

Figure 14 provides a nucleotide (SEQ ID NO:19) and an amino acid (SEQ ID NO:20) sequence of human HCDR-3.

Figure 15 provides a nucleotide (SEQ ID NO:21) and an amino acid (SEQ ID NO:22) sequence of human DDX21.

Figure 16 provides a nucleotide (SEQ ID NO:23) and an amino acid (SEQ ID NO:24) sequence of human ARK2.

Figure 17 provides a nucleotide (SEQ ID NO:25) and an amino acid (SEQ ID NO:26) sequence of human transmembrane 4 superfamily member 1.

Figure 18 provides a nucleotide (SEQ ID NO:27) and an amino acid (SEQ ID NO:28) sequence of human ERCC1.

Figure 19 provides an illustration of certain relevant domains of FANCA, including the aldehyde dehydrogenase cysteine active site, FKBP-type peptidyl-prolyl cis-trans isomerase signature 1 site, the PX site, and the peptidase S8 site.

Figure 20 illustrates cell tracker assay data demonstrating that GFP-fused FANCA is antiproliferative in A549 cancer cells. The FANCA construct is the functional hit isolated in the retroviral screen. Figure 20 top left illustrates fluorescence analysis of green fluorescent protein (GFP) infected A549.tTA control cells. Figure 20 top right illustrates cell tracker assay data from GFP infected A549.tTA control cells. Figure 20 lower left illustrates fluorescence analysis of FANCA infected A549.tTA cells. Figure 20 lower right illustrates cell tracker assay date from FANCA infected A549.tTA cells.

Figure 21 provides an illustration of certain relevant domains of DDX9, including the double stranded RNA binding motif, the DEAD/DEAH box helicase domain, the helicase conserved C terminal domain, and the GLN3 protein domain.

Figure 22 illustrates cell tracker assay data demonstrating that GFP-fused DDX9 is antiproliferative in A549 cancer cells. The DDX9 construct is the functional hit isolated in the retroviral screen. Figure 22 top left illustrates fluorescence analysis of green fluorescent protein (GFP) infected A549.tTA control cells. Figure 22 top right illustrates cell tracker assay data from GFP infected A549.tTA control cells. Figure 22 lower left illustrates fluorescence analysis of DDX9 infected A549.tTA cells. Figure 22 lower right illustrates cell tracker assay date from DDX9 infected A549.tTA cells.

Figure 23 provides an illustration of certain relevant domains of IGF1R, including the receptor L domain, the furin-like cysteine rich region, the fibronectin type II domain, the transmembrane domain, and the kinase domain.

Figure 24 illustrates cell tracker assay data demonstrating that GFP-fused IGF1R is antiproliferative in A549. The IGF1R construct is the functional hit isolated in the retroviral screen. Figure 24 top left illustrates fluorescence analysis of green fluorescent protein (GFP) infected A549.tTA control cells. Figure 24 top right illustrates cell tracker assay data from GFP infected A549.tTA control cells. Figure 24lower left illustrates

fluorescence analysis of IGF1R infected A549.tTA cells. Figure 24 lower right illustrates cell tracker assay date from IGF1R infected A549.tTA cells.

Figure 25 provides an illustration of the relevant domains of UBE2V1, including the ubiquitin conjugating enzyme domain.

Figure 26 illustrates cell tracker assay data demonstrating that GFP-fused UBE2V1 is antiproliferative in A549 cancer cells. The UBE2V1 construct is the functional hit isolated in the retroviral screen. Figure 26 top left illustrates fluorescence analysis of green fluorescent protein (GFP) infected A549.tTA control cells. Figure 26 top right illustrates cell tracker assay data from GFP infected A549.tTA control cells. Figure 26 lower left illustrates fluorescence analysis of UBE2V1 infected A549.tTA cells. Figure 26 top right illustrates cell tracker assay date from UBE2V1 infected A549.tTA cells.

Figure 27 shows four alternatively spliced UBE2V1 transcripts.

Figure 28 provides some cDNA sequence isolated from a cell tracker assay for cDNAs that regulate the cell cycle, *i.e.*, functional hits from the retroviral screen.

Figure 29 provides dominant negative mutants of BAP-1. Mutated residues are shown with arrows.

Figure 30 provides evidence that expression of Bap1 WT and protease mutants is antiproliferative in HeLa cells.

Figure 31 provides evidence tthat expression of Bap1 WT protein is antiproliferative in HeLa cells in the Celltracker assay.

Figure 32 provides evidence that expression of Bap1 protease mutants is slightly more antiproliferative than expression of Bap1 WT in H1299 cells.

Figure 33 provides evidence expression of Bap1 WT and Bap1 protease mutants is antiproliferative in H1299 cells in the Celltracker assay.

Figure 34 provides evidence that the Bap1 functional hit G32D8 is antiproliferative in HMEC cells.

Figure 35 provides evidence that the Bap1 functional hit G3-2D8 is antiproliferative in PrEC cells.

Figure 36 provides evidence that BAP1 specific siRNA has an antiproliferative effect on HeLa cells.

Figure 37 provides evidence that BAP1 specific siRNA induces G1 arrest in H1299 cells.

Figure 38 provides evidence that soluble GST-Bap1 protein can be expressed from SF9 cells. GST-Bap1 was produced using the baculovirus transfer vector pDEST20

along with the Bac-to-Bac baculovirus expression system (invitrogen). GST-Bap1(1) and GST-Bap1(2) refer to two different virus dilutions used for expression.

Figure 39 provides SDS-PAGE gels showing BAP-1 purification.

Figure 40 provides an example of a fluorogenic Ub cleavage assay. Aminomethyl-coumarin cleavage from a Ub C-terminus generates fluorescence emission in the solution-phase assay.

Figure 41 provides evidence that BAP1 is an active ubiquitin protease.

Figure 42 demonstrates the kinetics of UbAMC cleavage by BAP1. The K_m is 0.5 μM.

Figure 43 provides evidence that UbCHO acts as specific inhibitor of BAP1 protease activity.

Figure 44 demonstrates that the Np95 functional hit G1-2635 is antiproliferative in HMEC cells.

Figure 45 demonstrates that the Np95 functional hit G1-2635 is antiproliferative in PrEC cells.

Figure 46 demonstrates that NP95 specific siRNAs have an antiproliferative effect on PrECs.

Figure 47 demonstrates that NP95 specific siRNAs induce G1 arrest in HUVEC cells.

Figure 48 demonstrates Taqman analysis (real time PCR) of NP95 mRNA expression in samples obtained from patients with breast carcinoma. Normal and tumor tissue samples from the same patient were analyzed.

Figure 49 demonstrates Taqman analysis of NP95 mRNA expression in samples obtained from patients with lung carcinoma. Normal and tumor tissue samples from the same patient were analyzed.

Figure 50 demonstrates Taqman analysis of NP95 mRNA expression in samples obtained from patients with prostate adenocarcinoma. Normal and tumor tissue samples from the same patient were analyzed. All tumors were of acinar cell origin.

Figure 51 provides dominant negative mutants for Np95. The RING finger domain of the protein was mutated.

Figure 52 demonstrates that GFP-fused Np95 ring finger mutants are slightly more antiproliferative than GFP-fused Np95 WT in HCT116 cells.

Figure 53 demonstrates that no antiproliferative effects are observed for Np95 WT and ring finger mutant constructs in A549 cells.

Figure 54 demostrates that A549 Cells expressing GFP-Np95 ΔRing become sensitized to bleomycin treatment.

Figure 55 demonstrates that Np95 WT and RING finger mutant constructs are strongly antiproliferative in HMECs.

Figure 56 demonstrates that Np95 WT and RING finger mutant constructs are strongly antiproliferative in PrECs.

Figure 57 demonstrates that NP95-specific siRNAs are antiproliferative in H1299 cells.

Figure 58 provides a schematic of the biochemistry of ubiqutination. NP95 is believed to be an E3 protein.

Figure 59 demonstrates that GFP-Np95 exhibits E3 ubiquitin ligase activity.

Figure 60 demonstrates that the RING domain is required for GFP-Np95 ligase activity.

Figure 61 demonstrates that NP95 WT can be expressed and purified from SF9 cells.

Figure 62 provides a plate-based ubiquitin ligase assay. The assay is also described in WO 01/75145, herein incorporated by reference for all purposes.

Figure 63 demonstrates NP95 activity in the plate-based auto-ubiquitylation assay. Reactions contained 100 ng FI-Ub, 5 ng of E1 and, 20 ng of E2 per well. The Np95 controls contained 150 ng Np95. The E3 control contained 75 ng E3. The two data sets are results of duplicate assays.

DETAILED DESCRIPTION OF THE INVENTION

INTRODUCTION

The BAP-1 gene encodes a 90 kDa (729 aa) ubiquitin carboxy-terminal hydrolase (UCH). BAP-1 has a ubiquitin carboxy-terminal hydrolase domain and a DNA binding domain. (*See, e.g.*, Irminger-Finger *et al., Biol. Chem.* 380(2):117 (1999), Jensen *et al., Oncogene* 16(9):1097 (1998)), and Jensen *et al., Ann. N.Y. Acad. Sci.* 886:191 (1999)). UCH family members are 25-30 kDa proteins that are typically localized to the cytoplasm. UCH family members cleave ubiquitin from ubiquitin conjugated small substrates and are postulated to be involved in cotranslational processing of proubiquitin. BAP-1 in particular is postulated to play a role in: deubiquitination of histones leading to chromatin rearrangement, deubiquitination of multiple transcription factors, and hydrolysis of ubiquitin like proteins. (*See, e.g.*, Jensen *et al., Ann. N.Y. Acad. Sci.* 886:191 (1999)).

BAP-1 was identified as a BRCA1 associated protein which binds to the BRCA1 RING finger domain. (*See, e.g.*, Jensen *et al., Oncogene* 16(9):1097 (1998)). BAP-1 has been shown to enhance BRCA1 mediated inhibition of breast cancer cell proliferation and is therefore postulated to be a tumor suppressor. (*See, e.g.*, Jensen *et al., Oncogene* 16(9):1097 (1998)). However, direct BAP-1 involvement in cellular transformation, tumorigenesis, and anti-proliferative effects in tumor cells has never been demonstrated. Furthermore, the role of BAP-1 in cell cycle regulation has not yet been elucidated.

The present inventors identified human BAP-1 in a cDNA library screening assay. As shown in Figure 3, studies with BAP-1 show BAP-1 has an antiproliferative phenotype for tumor cells (using, e.g., GFP positivity and cell tracker assays). These functional studies, presented herein, demonstrate for the first time that inhibition of BAP-1 will inhibit tumor cell growth. In BAP-1 infected A549.tTA cells, fluorescence analysis indicates that BAP-1 may be localized to the cytoplasm.

The NP95 gene encodes a nuclear zinc finger protein which is associated with cellular proliferation (*see, e.g.*, Fujimori *et al. Mammalian Genome* 9:1032-1035 (1998). The NP95 open reading frame contains a potential ATP/GTP binding site, a zinc finger motif, a putative cyclin A/E cdk2 phosphorylation site, and a retinoblastoma binding motif (*see, e.g.*, Miura *et al. Exp. Cell Res.* 263:202-208 (2001). However, NP95 involvement in cellular transformation, tumorigenesis, and anti-proliferative effects in tumor cells has never been demonstrated. Furthermore, the role of NP95 in cell cycle regulation has not yet been elucidated.

As described below, the present inventors identified human NP95 in a cDNA library screening assay. As shown in Figure 6, studies with NP95 show NP95 has an antiproliferative phenotype (using, e.g., GFP positivity and cell tracker assays). These functional studies, presented herein, demonstrate for the first time that inhibition of NP95 will inhibit tumor cell growth. With cellular staining of NP95 infected A549.tTA cells, fluorescence analysis shows that NP95 is localized to the nucleus of NP95 infected cells.

BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, and ERCC1 encode proteins involved in modulation of the cell cycle in cancer cells.

As described below, the present inventors identified BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, and ERCC1 as modulators of the cell cycle in a cDNA library screening assay.

In one embodiment, as shown in Figure 20, studies with FANCA show FANCA has an antiproliferative phenotype (using, e.g., GFP positivity and cell tracker assays). Cellular staining of FANCA infected A549.tTA cells shows that FANCA may be localized to the cytoplasm.

In one embodiment, as shown in Figure 22, studies with DDX9 show DDX9 has an antiproliferative phenotype (using, e.g., GFP positivity and cell tracker assays). Cellular staining of DDX9 infected A549.tTA cancer cells shows that DDX9 may be localized to the cytoplasm.

In one embodiment, as shown in Figure 24, studies with IGF1R show IGF1R has an antiproliferative phenotype (using, e.g., GFP positivity and cell tracker assays). Cellular staining of IGF1R infected A549.tTA cancer cells shows that IGF1R is localized to the cytoplasm.

In one embodiment, as shown in Figure 26, studies with UBE2V1 show UBE2V1 has an antiproliferative phenotype (using, e.g., GFP positivity and cell tracker assays). These functional studies, presented herein, demonstrate for the first time that inhibition of FANCA, DDX9, IGF1R, and UBE2V1 will inhibit tumor cell growth. Cellular staining of UBE2V1 infected A549.tTA cancer cells shows that UBE2V1 may be localized to the cytoplasm.

The FANCA gene is approximately 80 kb and has been localized to chromosome 16q24.3 (*see, e.g.*, Pronk *et al., Nat. Genet.* 11:338-340 (1995); Foe *et al., Nat. Genet.* 14:320-323 (1996); Ianzano *et al., Genomics* 41:309-314 (1997); Joenje *et al., Am. J. Hum. Genet.* 61:940-944 (1997); and Kupfer *et al., Nat. Genet.* 17:487-490 (1997)). The N terminal region of FANCA encodes a putative peroxidase domain (*see* Ren & Youssoufian, *Mol. Gen . Metabol.* 72:54 (2001)). FANCA has been found to associate with BRG1, a component of SWI/SNF, a complex active in regulation of transcription (*see* Otsuki *et al., Hum. Mol. Genet.* 10(23):2651 (2001)). Assays such as enzymatic activity assays known to those of skill in the art can be used to identify modulators or FANCA, e.g., aldehyde dehydrogenase activity.

DDX9 encodes RNA helicase A and the identical protein nuclear DNA helicase II (*see, e.g.*, Lee & Hurwitz, *J. Biol. Chem.* 267:4398-4407 (1992); Lee *et al., J. Biol. Chem.* 268:13472-13478 (1993); Lee & Hurwitz, *J. Biol. Chem.* 268:16822-16830 (1993); Abdelhaleem *et al., J. Immunol.* 156:2026-2035 (1996); Zhang & Grosse, *J. Biol. Chem.* 272:11487-11494 (1997); Nakajima *et al., Cell* 90:1107-1112 (1997); Lee *et al., Proc. Nat'l Acad. Sci. USA* 95:13709-13713 (1998); Lee *et al., Somat. Cell Mol. Genet.* 25:33-39 (1999);

Imamura, *et al., Nuc. Acids Res.* 26(9):2063 (1998); and Zhang *et al., J. Cell. Sci.* 112:2693 (1999)). Vectors containing DNA encoding DDX9 complement yeast that have mutations in *prp8-1*, the yeast homolog of DDX9 (*see* Imamura *et al.*). Helicase assays known to those of skill in the art can be used, e.g., to identify modulators of DDX9.

IGF1R encodes a cell surface tyrosine kinase receptor and binds to IGF1 ligand (*see, e.g.*, Nakae *et al., Endocr. Rev.* 22(6):818 (2001); Flier *et al., Proc. Nat'l Acad. Sci. USA* 83:664-668 (1086); Francke *et al., Cold Spring Harb. Symp. Quant. Biol.* 51(Pt. 2):855-866 (1986); Ullrich *et al., EMBO J.* 5:2503-2512 (1986); Cooke *et al., Biochem. Biophys. Res. Commun.* 177:1113-1120 (1991); Abbott *et al., J. Biol. Chem.* 267:10759-10763 (1992); Werner *et al., Proc. Nat'l Acad. Sci. USA* 93:318-8323 (1996); Grant *et al., J. Clin. Endocrinol. Metab.* 83:3252-3257 (1998); and Butler & LeRoith, *Endocrinology* 142(5):1685 (2001)). Upon ligand binding, the receptor undergoes a conformational change which enables it to bind ATP, thereby increasing their kinase activity and modulate cell proliferation (*see* Nakae *et al.*). IGF1R deficient mice develop cell proliferation disorders, including muscle hypoplasia due to decreased cell numbers; IGF1R null mice develop cell proliferation disorders including dwarfism (*Id.*). Overexpression of IGF1R has been linked to increased radioresistance of breast cancer cells (*see* Macaulay *et al., Oncogene* 22(6):4029 (2001)). Ligand binding assays, autophosphorylation assays, kinase assays, and signal transduction assays known to those of skill in the art can be used, e.g., to identify modulators of IGF1R.

UBE2V1 encodes a protein that has been show to play a role in cell cycle regulation (*see, e.g.*, Rothofsky *et al., Gene* 195:141-149 (1997); Sancho *et al., Mol. Cell. Biol.* 18:576-589 (1998); Ma *et al., Oncogene* 17:1321-1326 (1998); Hofmann & Pickart, *Cell* 96:645-653 (1999); Deng *et al., Cell* 103:351-361 (2000); and Thomson *et al., Genome Res.* 10:1743-1756 (2000)). Constitutive expression of exogenous UBE2V1 inhibits the capacity of colorectal adenocarcinoma cells to differentiate upon confluence and inhibits the mitotic kinase cdk1, thereby inducing the cells to arrest at the G_2-M phase of the cell cycle (*see,* Sancho *et al., Mol. Cell. Biol.* 18(1):576 (1998) and Stubbs *et al., Am. J. Path.* 154(5):1335 (1999)). UBE2V1 has four alternatively spliced transcripts that encode proteins with the conserved Ubc domain of E2 enzymes and unique N-terminal sequence (see Figure 21). Ubiquitination assays, e.g., ubiquitin ligase assays, known to those of skill in the art, can be used to identify modulators of UBE2V1.

Aldehyde dehydrogenases form a superfamily of NADP+ dependent enzymes that are involved in several distinct metabolic pathways (*see* Vasilou *et al., Chem. Biol.*

Interact. 129(1-2):1 (2000); Vasilou & Pappa, *Pharmacology* 61(3):192 (2000); Hsu *et al., Proc. Nat'l Acad. Sci USA* 82:3771-3775 (1985); Raghunathan *et al., Genomics* 2:267-269 (1988); Hsu *et al., Genomics* 5:857-865 (1989); Pereira *et al., Biochem. Biophys. Res. Comm.* 175:831-838 (1991); Zheng *et al., Alcohol. Clin. Exp. Res.* 17:828-838 (1993); Kathmann & Lipsky, *Biochem. Biophys. Res. Commun.* 236:527-531 (1997)). Loss of function mutations in aldehyde dehydrogenase genes lead to metabolic disorders including Sjögren-Larsson syndrome, type II hyperprolinemia, and 4-hydroxybutyric aciduria. Enzyme activity assays known to those of skill in the art can be used to identify modulators of aldehyde dehydrogenase.

Pyruvate kinase plays a key role in the metabolic pathway of glycolysis. Pyruvate kinase is typically a tetramer of 4 identical 500-600 amino acid subunits (*see* Wang *et al., Blood* 98(10):3113 (2001)). Pyruvate kinase deficiency is a leading cause of hereditary nonspherocytic hemolytic anemia (*see* Beutler & Gelbart, *Blood* 95(11):3585 (2000)). Pyruvate kinase deficiency has been linked to metabolic disorders, including the Crabtree effect in which proliferating cells exhibit decreased respiratory activity during glucose utilization (*see* Melo *et al., Cell Biochem. Func.* 16:99 (1998)). Kinase assays known to those of skill in the art can be used to identify modulators of pyruvate kinase.

G6PD encodes a key metabolic enzyme that catalyzes the oxidation of glucose-6-phosphate to 6-phosphogluconolactone (*see* Ho *et al., Free Rad. Biol. Med.* 29(2):156 (2000)). G6PD is linked to neonatal jaundice, drug induced hemolytic crisis, infection induced hemolytic crisis, favism, and nonspherocytic hemolytic anemia. (*Id.*). G6PD deficient cells exhibit increased doubling time, and premature senescence by arresting in G_1 phase (*Id.*). It has also been reported that women with G6PD deficiency have a decreased risk of breast cancer (*see* Di Monco *et al. Br. J. Canc.* 75(4):589 (1997)). Enzyme activity assays known to those of skill in the art can be used to identify modulators of G6PD.

HCDR-3, also called proliferation associated 2G4, encodes a protease. Protease assays known to those of skill in the art can be used to identify modulators of HCDR-3.

DDX21 encodes a RNA helicase II. DDX21 hydrolyzes ATP and dATP in the presence of RNA, unwinds dsRNA in the 5' to 3' direction, and folds ssRNA (*see*, Valdez, *Eur. J. Biochem.* 267:6395 (2000)). Autoantibodies to DDX21 have been found in patients with connective tissue diseases, including watermelon stomach disease (*see* Ou *et al., Exp. Cell Res.* 247:389 (1999) and Valdez *et al. Nuc. Acids. Res.*, 24(7):1220 (1996)). Helicase assays known to those of skill in the art can be used to identify modulators of DDX21.

ARK2 encodes a serine threonine kinase which is postulated to play a role in mitosis (*see* Descamps & Prigent, *Sci. STKE* 173:1 (2001)). Specifically, ARK2 has been shown to accumulate in the midbodies during mitosis (*see* Shindo *et al.*, *Biochem. Biophys. Res. Commun.* 244(1):285 (1998)). ARK2 deficient cells have also been shown to exhibit cytokinesis defects (Descamps & Prigent,). Kinase assays known to those of skill in the art can be used to identify modulators of ARK2.

Transmembrane 4 superfamily 1 is a member of a family of cell surface molecules with four hydrophobic domains, are widely expressed, and have roles in diverse cellular functions, including cell proliferation, cell signaling, cell motility, and tumor metastasis (*see* Class *et al., J. Biol. Chem.* 276(11):7974 (2001) and Zhang *et al., J. Biol. Chem.* 276(27):25005 (2001)). Studies in knockout mice lacking transmembrane 4 superfamily 1 proteins have shown that the protein is a potent regulator of lymphocyte proliferation (*see* Miyazaki *et al., EMBO* 16(14):4217 (1997)). Signal transduction assays and cellular proliferation assays known to those of skill in the art can be used to identify modulators of transmembrane 4 superfamily 1.

ERCC1 encodes a nucleotide excision repair gene (*see* Nunez *et al., FASEB J.* 14:1073 (2000)). ERCC1 knockout mice have hepatocytes that are arrested in G_2 phase and have reduced DNA replication and binucleation (*Id.*). Immortalized embryonic fibroblasts from ERCC1 deficient mice exhibit increased genome instability (*see* Melton *et al., J.Cell Sci.* 111:395 (1998)). ERCC1 knockout mice also severely runted and have a greatly shortened lifespan when compared to normal mice (*see* Weeda *et al., Curr. Biol.* 7:427 (1997)). DNA repair and endonuclease assays known to those of skill in the art can be used to identify modulators of ERCC1.

Thus, BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DEAD/H box polypeptide 21 (DDX21), serine threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, and ERCC1 can conveniently be used to identify agents that modulate the cell cycle.

BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, and ERCC1 therefore represent drug targets for compounds that suppress or activate cellular proliferation in tumor cells, or cause cell cycle arrest, cause release from cell cycle arrest,

activate apoptosis, increase sensitivity to chemotherapeutic (adjuvant) reagents, and decrease toxicity of chemotherapeutic reagents. Agents identified in these assays, including small organic molecules, peptides, cyclic peptides, nucleic acids, antibodies, antisense nucleic acids, RNAi, and ribozymes, that modulate cell cycle regulation and cellular proliferation via modulation of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1, can be used to treat diseases related to cellular proliferation, such as cancer. In particular, inhibitors of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 are useful for inhibition of cancer and tumor cell growth. BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 modulators can also be used to modulate the sensitivity of cells to chemotherapeutic agents, such as bleomycin, etoposide, taxol, and other agents known to those of skill in the art BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 modulators can also be used to decrease toxicity of such chemotherapeutic reagents.

In one embodiment, enzymatic assays, including ubiquitin hydrolase assays, ubiquitin ligase assays, kinase or autophosphorylation assays, RNA helicase assays, pyruvate kinase assays, aldehyde dehydrogenase assays, and glucose-6-phosphate dehydrogenase assays using BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 can be used to identify modulators of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 activity, or to identify proteins that bind to BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1, e.g., BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 substrates. Full length wild type BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1, mutant BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1

Such modulators are useful for treating cancers, such as melanoma, breast, ovarian, lung, gastrointestinal and colon, prostate, and leukemia and lymphomas, e.g., multiple myeloma. In addition, such modulators are useful for treating noncancerous disease states caused by pathologically proliferating cells such as thyroid hyperplasia (Grave's disease), psoriasis, benign prostatic hypertrophy, neurofibromas, atherosclerosis, restenosis, and other vasoproliferative disease.

Cell proliferation assays described herein reveal for the first time that expression of a nucleic acid molecule encoding the above described cell cycle regulatory proteins (*i.e.*, BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1) exerted a negative effect on cellular proliferation. Without wishing to be bound by theory, it appears that the cell cycle regulatory proteins (*i.e.*, BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1) or fragments of those proteins, peptides derived from the proteins, or peptides and inhibitory DNA or RNA molecules derived from DNA encoding the proteins, provide an anti-proliferative phenotype. Thus, in addition to their use in screens for modulators of the cell cycle, the cell cycle regulatory proteins (*i.e.*, BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1) or fragments of those proteins, peptides derived from the proteins, or peptides and inhibitory DNA or RNA molecules derived from DNA encoding the proteins, can also be used as therapeutics for treatment of cancers, such as melanoma, breast, ovarian, lung, gastrointestinal and colon, prostate, and leukemia and lymphomas, e.g., multiple myeloma. In addition, such modulators are useful for treating noncancerous disease states caused by pathologically proliferating cells such as thyroid hyperplasia (Grave's disease), psoriasis, benign prostatic hypertrophy, neurofibromas, atherosclerosis, restenosis, and other vasoproliferative disease.

DEFINITIONS

By "disorder associated with cellular proliferation" or "disease associated with cellular proliferation" herein is meant a disease state which is marked by either an excess or a deficit of cellular proliferation or apoptosis. Such disorders associated with increased cellular proliferation include, but are not limited to, cancer and non-cancerous pathological proliferation. BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase,

pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein levels ;or levels of a nucleic acid encoding BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 can be determined and used for diagnostic or prognostic testing of subjects believed to have a disorder or disease associated with cellular proliferation.

The terms "BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1" or a nucleic acid encoding "BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1" refer to nucleic acids and polypeptide polymorphic variants, alleles, mutants, and interspecies homologs that: (1) have an amino acid sequence that has greater than about 60% amino acid sequence identity, 65%, 70%, 75%, 80%, 85%, 90%, preferably 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% or greater amino acid sequence identity, preferably over a region of over a region of at least about 25, 50, 100, 200, 500, 1000, or more amino acids, to an amino acid sequence encoded by a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 nucleic acid (for a human BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 nucleic acid sequence, *see, e.g.*, Figures 1, 4, and 7-18, SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 or Accession number NM_004656, AF274048, NM_000135, NM_000875, NM_030588, NM_003349, NM_000689, XM_037768.1, XM_049337.1, XM_030607.1, XM_027538.1, BC008442, XM_049047.1, and XM_052326.1) or amino acid sequence of a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein (for a human BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein sequence, *see, e.g.*, Figures 1, 4, and 7-18, SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 or Accession numbers NM_004656, AF274048, NM_000135, NM_000875, NM_030588, NM_003349, NM_000689, XM_037768.1, XM_049337.1, XM_030607.1, XM_027538.1, BC008442, XM_049047.1, and XM_052326.1 (see also NP_004647, AAK55744.1, NP_000126, NP_000866, NP_085077, NP_003340, NP_000680, and

NP_S30038)); (2) bind to antibodies, e.g., polyclonal antibodies, raised against an immunogen comprising an amino acid sequence of a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein, and conservatively modified variants thereof; (3) specifically hybridize under stringent hybridization conditions to an anti-sense strand corresponding to a nucleic acid sequence encoding a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein, and conservatively modified variants thereof; (4) have a nucleic acid sequence that has greater than about 95%, preferably greater than about 96%, 97%, 98%, 99%, or higher nucleotide sequence identity, preferably over a region of at least about 25, 50, 100, 200, 500, 1000, or more nucleotides, to a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 nucleic acid or a nucleic acid encoding the enzymatic domain. Preferably the enzymatic domain has greater than 96%, 97%, 98%, or 99% amino acid identity to the human BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 enzymatic domain of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, or 28. A polynucleotide or polypeptide sequence is typically from a mammal including, but not limited to, primate, e.g., human; rodent, e.g., rat, mouse, hamster; cow, pig, horse, sheep, or any mammal. The nucleic acids and proteins of the invention include both naturally occurring or recombinant molecules.

The phrase "functional effects" in the context of assays for testing compounds that modulate activity of a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, OR ERCC1 protein includes the determination of a parameter that is indirectly or directly under the influence of a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1, e.g., a phenotypic or chemical effect, such as the ability to increase or decrease cellular proliferation, apoptosis, cell cycle arrest, or enzymatic activity, or e.g., a physical effect such as ligand binding or inhibition of ligand binding. A functional effect therefore includes ligand binding activity, the ability of cells to proliferate, apoptosis, and enzyme activity. "Functional effects" include *in vitro*, *in vivo*, and *ex vivo* activities.

By "determining the functional effect" is meant assaying for a compound that increases or decreases a parameter that is indirectly or directly under the influence of a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein, e.g., measuring physical and chemical or phenotypic effects. Such functional effects can be measured by any means known to those skilled in the art, e.g., changes in spectroscopic characteristics (e.g., fluorescence, absorbance, refractive index); hydrodynamic (e.g., shape); chromatographic; or solubility properties for the protein; measuring inducible markers or transcriptional activation of the protein; measuring binding activity or binding assays, e.g. binding to antibodies; measuring changes in ligand or substrate binding activity; measuring receptor binding, measuring receptor cross-linking or other intracellular response to receptor binding; measuring cellular proliferation; measuring cell morphology, e.g., spindle formation or chromosome formation; measuring phosphorylated proteins such as histone H3 using antibodies; measuring apoptosis; measuring cell surface marker expression; measurement of changes in protein levels for BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1-associated sequences; measurement of RNA stability; identification of downstream or reporter gene expression (CAT, luciferase, β-gal, GFP and the like), e.g., via chemiluminescence, fluorescence, colorimetric reactions, antibody binding, and inducible markers. In one embodiment, the function effect is determined using an *in vitro* ubiquitin ligase assay or a ubiquitin conjugation assay as described in Examples 2 and 3 of WO 01/17145, using recombinant ubiquitin and ubiquitin-like molecules, E1, E2, and E3 molecules of choice, e.g., NP95. In a preferred embodiment, a substrate free, auto E3 ubiquitin ligase assay can be used in the methods of the invention (*see, e.g.*, WO 01/75145).

"Inhibitors", "activators", and "modulators" of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 polynucleotide and polypeptide sequences are used to refer to activating, inhibitory, or modulating molecules identified using *in vitro* and *in vivo* assays of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 polynucleotide and polypeptide sequences. Inhibitors are compounds that, e.g., bind to, partially or totally block activity, decrease, prevent, delay activation, inactivate, desensitize, or down regulate the activity or

expression of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 proteins, e.g., antagonists. "Activators" are compounds that increase, open, activate, facilitate, enhance activation, sensitize, agonize, or up regulate BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein activity, e.g., agonists. Inhibitors, activators, or modulators also include genetically modified versions of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 proteins, e.g., versions with altered activity, as well as naturally occurring and synthetic ligands, antagonists, agonists, antibodies, peptides, cyclic peptides, nucleic acids, antisense molecules, RNAi molecules, ribozymes, small chemical molecules and the like. Such assays for inhibitors and activators include, e.g., expressing BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein *in vitro*, in cells, or cell membranes, applying putative modulator compounds, and then determining the functional effects on activity, as described above.

Samples or assays comprising BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 proteins that are treated with a potential activator, inhibitor, or modulator are compared to control samples without the inhibitor, activator, or modulator to examine the extent of inhibition. Control samples (untreated with inhibitors) are assigned a relative protein activity value of 100%. Inhibition of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 is achieved when the activity value relative to the control is about 80%, preferably 50%, more preferably 25-0%. Activation of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 is achieved when the activity value relative to the control (untreated with activators) is 110%, more preferably 150%, more preferably 200-500% (i.e., two to five fold higher relative to the control), more preferably 1000-3000% higher.

The term "test compound" or "drug candidate" or "modulator" or grammatical equivalents as used herein describes any molecule, either naturally occurring or synthetic, e.g., protein, oligopeptide (e.g., from about 5 to about 25 amino acids in length, preferably

from about 10 to 20 or 12 to 18 amino acids in length, preferably 12, 15, or 18 amino acids in length), small organic molecule, polysaccharide, lipid, fatty acid, polynucleotide, oligonucleotide, etc., to be tested for the capacity to directly or indirectly modulation tumor cell proliferation. The test compound can be in the form of a library of test compounds, such as a combinatorial or randomized library that provides a sufficient range of diversity. Test compounds are optionally linked to a fusion partner, e.g., targeting compounds, rescue compounds, dimerization compounds, stabilizing compounds, addressable compounds, and other functional moieties. Conventionally, new chemical entities with useful properties are generated by identifying a test compound (called a "lead compound") with some desirable property or activity, e.g., inhibiting activity, creating variants of the lead compound, and evaluating the property and activity of those variant compounds. Often, high throughput screening (HTS) methods are employed for such an analysis.

A "small organic molecule" refers to an organic molecule, either naturally occurring or synthetic, that has a molecular weight of more than about 50 daltons and less than about 2500 daltons, preferably less than about 2000 daltons, preferably between about 100 to about 1000 daltons, more preferably between about 200 to about 500 daltons.

"RNAi molecule" or an "siRNA" refers to a nucleic acid that forms a double stranded RNA, which double stranded RNA has the ability to reduce or inhibit expression of a gene or target gene when the siRNA expressed in the same cell as the gene or target gene. "siRNA" thus refers to the double stranded RNA formed by the complementary strands. The complementary portions of the siRNA that hybridize to form the double stranded molecule typically have substantial or complete identity. In one embodiment, an siRNA refers to a nucleic acid that has substantial or complete identity to a target gene and forms a double stranded siRNA. The sequence of the siRNA can correspond to the full length target gene, or a subsequence thereof. Typically, the siRNA is at least about 15-50 nucleotides in length (e.g., each complementary sequence of the double stranded siRNA is 15-50 nucleotides in length, and the double stranded siRNA is about 15-50 base pairs in length, preferable about preferably about 20-30 base nucleotides, preferably about 20-25 nucleotides in length, *e.g.*, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides in length.

"Ubiquitin ligation pathway or component" refers to ubiquitin and ubiquitin-like molecules (see Figure 58), and E1, E2, and E3 proteins and their substrates, which are involved in the ubiquitination process (*see, e.g.*, Weissman, *Nature Reviews* 2:169-178 (2001); *see also* WO 01/75145)).

"Biological sample" include sections of tissues such as biopsy and autopsy samples, and frozen sections taken for histologic purposes. Such samples include blood, sputum, tissue, cultured cells, e.g., primary cultures, explants, and transformed cells, stool, urine, etc. A biological sample is typically obtained from a eukaryotic organism, most preferably a mammal such as a primate e.g., chimpanzee or human; cow; dog; cat; a rodent, e.g., guinea pig, rat, mouse; rabbit; or a bird; reptile; or fish.

The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (i.e., about 60% identity, preferably 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or higher identity over a specified region (e.g., nucleotide sequence SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 or amino acid sequence SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28), when compared and aligned for maximum correspondence over a comparison window or designated region) as measured using a BLAST or BLAST 2.0 sequence comparison algorithms with default parameters described below, or by manual alignment and visual inspection. Such sequences are then said to be "substantially identical." This definition also refers to, or may be applied to, the compliment of a test sequence. The definition also includes sequences that have deletions and/or additions, as well as those that have substitutions. As described below, the preferred algorithms can account for gaps and the like. Preferably, identity exists over a region that is at least about 25 amino acids or nucleotides in length, or more preferably over a region that is 50-100 amino acids or nucleotides in length.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Preferably, default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences

for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection (*see, e.g., Current Protocols in Molecular Biology* (Ausubel *et al.*, eds. 1995 supplement)).

A preferred example of algorithm that is suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al., Nuc. Acids Res.* 25:3389-3402 (1977) and Altschul *et al., J. Mol. Biol.* 215:403-410 (1990), respectively. BLAST and BLAST 2.0 are used, with the parameters described herein, to determine percent sequence identity for the nucleic acids and proteins of the invention. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al., supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915

(1989)) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

"Nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. The term encompasses nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, peptide-nucleic acids (PNAs).

Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* 260:2605-2608 (1985); Rossolini *et al.*, *Mol. Cell. Probes* 8:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, mRNA, oligonucleotide, and polynucleotide.

A particular nucleic acid sequence also implicitly encompasses "splice variants." Similarly, a particular protein encoded by a nucleic acid implicitly encompasses any protein encoded by a splice variant of that nucleic acid. "Splice variants," as the name suggests, are products of alternative splicing of a gene. After transcription, an initial nucleic acid transcript may be spliced such that different (alternate) nucleic acid splice products encode different polypeptides. Mechanisms for the production of splice variants vary, but include alternate splicing of exons. Alternate polypeptides derived from the same nucleic acid by read-through transcription are also encompassed by this definition. Any products of a splicing reaction, including recombinant forms of the splice products, are included in this definition. An example of potassium channel splice variants is discussed in Leicher, *et al.*, *J. Biol. Chem.* 273(52):35095-35101 (1998).

The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding

naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymer.

The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, γ-carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

"Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a

polypeptide is implicit in each described sequence with respect to the expression product, but not with respect to actual probe sequences.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.

The following eight groups each contain amino acids that are conservative substitutions for one another: 1) Alanine (A), Glycine (G); 2) Aspartic acid (D), Glutamic acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W); 7) Serine (S), Threonine (T); and 8) Cysteine (C), Methionine (M) (*see, e.g.*, Creighton, *Proteins* (1984)).

Macromolecular structures such as polypeptide structures can be described in terms of various levels of organization. For a general discussion of this organization, *see, e.g.*, Alberts *et al.*, *Molecular Biology of the Cell* (3rd ed., 1994) and Cantor and Schimmel, *Biophysical Chemistry Part I: The Conformation of Biological Macromolecules* (1980). "Primary structure" refers to the amino acid sequence of a particular peptide. "Secondary structure" refers to locally ordered, three dimensional structures within a polypeptide. These structures are commonly known as domains, e.g., enzymatic domains, extracellular domains, transmembrane domains, pore domains, and cytoplasmic tail domains. Domains are portions of a polypeptide that form a compact unit of the polypeptide and are typically 15 to 350 amino acids long. Exemplary domains include domains with enzymatic activity, e.g., a kinase domain. Typical domains are made up of sections of lesser organization such as stretches of β-sheet and α-helices. "Tertiary structure" refers to the complete three dimensional structure of a polypeptide monomer. "Quaternary structure" refers to the three dimensional structure formed by the noncovalent association of independent tertiary units. Anisotropic terms are also known as energy terms.

A "label" or a "detectable moiety" is a composition detectable by spectroscopic, photochemical, biochemical, immunochemical, chemical, or other physical means. For example, useful labels include ^{32}P, fluorescent dyes, electron-dense reagents,

enzymes (e.g., as commonly used in an ELISA), biotin, digoxigenin, or haptens and proteins which can be made detectable, e.g., by incorporating a radiolabel into the peptide or used to detect antibodies specifically reactive with the peptide.

The term "recombinant" when used with reference, e.g., to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all.

The term "heterologous" when used with reference to portions of a nucleic acid indicates that the nucleic acid comprises two or more subsequences that are not found in the same relationship to each other in nature. For instance, the nucleic acid is typically recombinantly produced, having two or more sequences from unrelated genes arranged to make a new functional nucleic acid, e.g., a promoter from one source and a coding region from another source. Similarly, a heterologous protein indicates that the protein comprises two or more subsequences that are not found in the same relationship to each other in nature (e.g., a fusion protein).

The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acids, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. The T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m, 50% of the probes are occupied at equilibrium). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, preferably 10 times background hybridization. Exemplary stringent hybridization conditions can be as following: 50% formamide, 5x SSC, and 1% SDS,

incubating at 42°C, or, 5x SSC, 1% SDS, incubating at 65°C, with wash in 0.2x SSC, and 0.1% SDS at 65°C.

Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cases, the nucleic acids typically hybridize under moderately stringent hybridization conditions. Exemplary "moderately stringent hybridization conditions" include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 1X SSC at 45°C. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency. Additional guidelines for determining hybridization parameters are provided in numerous reference, e.g., and *Current Protocols in Molecular Biology*, ed. Ausubel, *et al.*

For PCR, a temperature of about 36°C is typical for low stringency amplification, although annealing temperatures may vary between about 32°C and 48°C depending on primer length. For high stringency PCR amplification, a temperature of about 62°C is typical, although high stringency annealing temperatures can range from about 50°C to about 65°C, depending on the primer length and specificity. Typical cycle conditions for both high and low stringency amplifications include a denaturation phase of 90°C - 95°C for 30 sec - 2 min., an annealing phase lasting 30 sec. - 2 min., and an extension phase of about 72°C for 1 - 2 min. Protocols and guidelines for low and high stringency amplification reactions are provided, e.g., in Innis *et al.* (1990) *PCR Protocols, A Guide to Methods and Applications*, Academic Press, Inc. N.Y.).

"Antibody" refers to a polypeptide comprising a framework region from an immunoglobulin gene or fragments thereof that specifically binds and recognizes an antigen. The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon, and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Light chains are classified as either kappa or lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively. Typically, the antigen-binding region of an antibody will be most critical in specificity and affinity of binding.

An exemplary immunoglobulin (antibody) structural unit comprises a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair

having one "light" (about 25 kD) and one "heavy" chain (about 50-70 kD). The N-terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms variable light chain (V_L) and variable heavy chain (V_H) refer to these light and heavy chains respectively.

Antibodies exist, e.g., as intact immunoglobulins or as a number of well-characterized fragments produced by digestion with various peptidases. Thus, for example, pepsin digests an antibody below the disulfide linkages in the hinge region to produce F(ab)'$_2$, a dimer of Fab which itself is a light chain joined to V_H-C_H1 by a disulfide bond. The F(ab)'$_2$ may be reduced under mild conditions to break the disulfide linkage in the hinge region, thereby converting the F(ab)'$_2$ dimer into an Fab' monomer. The Fab' monomer is essentially Fab with part of the hinge region (*see Fundamental Immunology* (Paul ed., 3d ed. 1993). While various antibody fragments are defined in terms of the digestion of an intact antibody, one of skill will appreciate that such fragments may be synthesized *de novo* either chemically or by using recombinant DNA methodology. Thus, the term antibody, as used herein, also includes antibody fragments either produced by the modification of whole antibodies, or those synthesized *de novo* using recombinant DNA methodologies (e.g., single chain Fv) or those identified using phage display libraries (*see, e.g.*, McCafferty *et al.*, *Nature* 348:552-554 (1990))

For preparation of antibodies, e.g., recombinant, monoclonal, or polyclonal antibodies, many technique known in the art can be used (*see, e.g.*, Kohler & Milstein, *Nature* 256:495-497 (1975); Kozbor *et al.*, *Immunology Today* 4: 72 (1983); Cole *et al.*, pp. 77-96 in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc. (1985); Coligan, *Current Protocols in Immunology* (1991); Harlow & Lane, *Antibodies, A Laboratory Manual* (1988); and Goding, *Monoclonal Antibodies: Principles and Practice* (2d ed. 1986)). The genes encoding the heavy and light chains of an antibody of interest can be cloned from a cell, e.g., the genes encoding a monoclonal antibody can be cloned from a hybridoma and used to produce a recombinant monoclonal antibody. Gene libraries encoding heavy and light chains of monoclonal antibodies can also be made from hybridoma or plasma cells. Random combinations of the heavy and light chain gene products generate a large pool of antibodies with different antigenic specificity (*see, e.g.*, Kuby, *Immunology* (3rd ed. 1997)). Techniques for the production of single chain antibodies or recombinant antibodies (U.S. Patent 4,946,778, U.S. Patent No. 4,816,567) can be adapted to produce antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms such as other mammals, may be used to express humanized or human antibodies (*see, e.g.*, U.S. Patent

Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, Marks *et al.*, *Bio/Technology* 10:779-783 (1992); Lonberg *et al.*, *Nature* 368:856-859 (1994); Morrison, *Nature* 368:812-13 (1994); Fishwild *et al.*, *Nature Biotechnology* 14:845-51 (1996); Neuberger, *Nature Biotechnology* 14:826 (1996); and Lonberg & Huszar, *Intern. Rev. Immunol.* 13:65-93 (1995)). Alternatively, phage display technology can be used to identify antibodies and heteromeric Fab fragments that specifically bind to selected antigens (*see, e.g.*, McCafferty *et al.*, *Nature* 348:552-554 (1990); Marks *et al.*, *Biotechnology* 10:779-783 (1992)). Antibodies can also be made bispecific, i.e., able to recognize two different antigens (*see, e.g.*, WO 93/08829, Traunecker *et al.*, *EMBO J.* 10:3655-3659 (1991); and Suresh *et al.*, *Methods in Enzymology* 121:210 (1986)). Antibodies can also be heteroconjugates, e.g., two covalently joined antibodies, or immunotoxins (*see, e.g.*, U.S. Patent No. 4,676,980 , WO 91/00360; WO 92/200373; and EP 03089).

Methods for humanizing or primatizing non-human antibodies are well known in the art. Generally, a humanized antibody has one or more amino acid residues introduced into it from a source which is non-human. These non-human amino acid residues are often referred to as import residues, which are typically taken from an import variable domain. Humanization can be essentially performed following the method of Winter and co-workers (*see, e.g.*, Jones *et al.*, *Nature* 321:522-525 (1986); Riechmann *et al.*, *Nature* 332:323-327 (1988); Verhoeyen *et al.*, *Science* 239:1534-1536 (1988) and Presta, *Curr. Op. Struct. Biol.* 2:593-596 (1992)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. Accordingly, such humanized antibodies are chimeric antibodies (U.S. Patent No. 4,816,567), wherein substantially less than an intact human variable domain has been substituted by the corresponding sequence from a non-human species. In practice, humanized antibodies are typically human antibodies in which some CDR residues and possibly some FR residues are substituted by residues from analogous sites in rodent antibodies.

A "chimeric antibody" is an antibody molecule in which (a) the constant region, or a portion thereof, is altered, replaced or exchanged so that the antigen binding site (variable region) is linked to a constant region of a different or altered class, effector function and/or species, or an entirely different molecule which confers new properties to the chimeric antibody, e.g., an enzyme, toxin, hormone, growth factor, drug, etc.; or (b) the variable region, or a portion thereof, is altered, replaced or exchanged with a variable region having a different or altered antigen specificity.

In one embodiment, the antibody is conjugated to an "effector" moiety. The effector moiety can be any number of molecules, including labeling moieties such as radioactive labels or fluorescent labels, or can be a therapeutic moiety. In one aspect the antibody modulates the activity of the protein.

The phrase "specifically (or selectively) binds" to an antibody or "specifically (or selectively) immunoreactive with," when referring to a protein or peptide, refers to a binding reaction that is determinative of the presence of the protein, often in a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein at least two times the background and more typically more than 10 to 100 times background. Specific binding to an antibody under such conditions requires an antibody that is selected for its specificity for a particular protein. For example, polyclonal antibodies raised to a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein, polymorphic variants, alleles, orthologs, and conservatively modified variants, or splice variants, or portions thereof, can be selected to obtain only those polyclonal antibodies that are specifically immunoreactive with BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 proteins and not with other proteins. This selection may be achieved by subtracting out antibodies that cross-react with other molecules. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select antibodies specifically immunoreactive with a protein (*see, e.g.*, Harlow & Lane, *Antibodies, A Laboratory Manual* (1988) for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity).

By "therapeutically effective dose" herein is meant a dose that produces effects for which it is administered. The exact dose will depend on the purpose of the treatment, and will be ascertainable by one skilled in the art using known techniques (*see, e.g.*, Lieberman, *Pharmaceutical Dosage Forms* (vols. 1-3, 1992); Lloyd, *The Art, Science and Technology of Pharmaceutical Compounding* (1999); and Pickar, *Dosage Calculations* (1999)).

ASSAYS FOR PROTEINS THAT MODULATE CELLULAR PROLIFERATION

High throughput functional genomics assays can be used to identify modulators of cellular proliferation. Such assays can monitor changes in cell surface marker expression, proliferation and differentiation, and apoptosis, using either cell lines or primary cells. Typically, the cells are contacted with a cDNA or a random peptide library (encoded by nucleic acids). In one embodiment, the peptides are cyclic or circular. The cDNA library can comprise sense, antisense, full length, and truncated cDNAs. The peptide library is encoded by nucleic acids. The effect of the cDNA or peptide library on the phenotype of cellular proliferation is then monitored, using an assay as described above. The effect of the cDNA or peptide can be validated and distinguished from somatic mutations, using, e.g., regulatable expression of the nucleic acid such as expression from a tetracycline promoter. cDNAs and nucleic acids encoding peptides can be rescued using techniques known to those of skill in the art, e.g., using a sequence tag.

Proteins interacting with the peptide or with the protein encoded by the cDNA (e.g., BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1) can be isolated using a yeast two-hybrid system, mammalian two hybrid system, immunoprecipitation or affinity chromatography of complexed proteins followed by mass spectrometry, or phage display screen, etc. Targets so identified can be further used as bait in these assays to identify additional members of the cellular proliferation pathway, which members are also targets for drug development (*see, e.g.*, Fields *et al.*, *Nature* 340:245 (1989); Vasavada *et al.*, *Proc. Nat'l Acad. Sci. USA* 88:10686 (1991); Fearon *et al.*, *Proc. Nat'l Acad. Sci. USA* 89:7958 (1992); Dang *et al.*, *Mol. Cell. Biol.* 11:954 (1991); Chien *et al.*, *Proc. Nat'l Acad. Sci. USA* 9578 (1991); and U.S. Patent Nos. 5,283,173, 5,667,973, 5,468,614, 5,525,490, and 5,637,463).

Suitable cell lines include A549, HeLa, Colo205, H1299, MCF7, MDA-MB-231, PC3, HMEC, PrEC. Cell surface markers can be assayed using fluorescently labeled antibodies and FACS. Cell proliferation can be measured using ^{3}H-thymidine incorporation, cell count by dye inclusion, MTT assay, BrdU incorporation, Cell Tracker assay, . Apoptosis can be measured using dye inclusion, or by assaying for DNA laddering, increases in intracellular calcium, or caspare activation. Growth factor production can be measured using an immunoassay such as ELISA.

cDNA libraries are made from any suitable source. Libraries encoding random peptides are made according to techniques well known to those of skill in the art (*see,*

e.g., U.S. Patent No. 6,153,380, 6,114,111, and 6,180,343). Any suitable vector can be used for the cDNA and peptide libraries, including, e.g., retroviral vectors.

ISOLATION OF NUCLEIC ACIDS ENCODING BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, ALDEHYDE DEHYDROGENASE, PYRUVATE KINASE, G6PD, HCDR-3, DDX21, ARK2, TRANSMEMBRANE 4 SUPERFAMILY MEMBER 1, OR ERCC1 FAMILY MEMBERS

This invention relies on routine techniques in the field of recombinant genetics. Basic texts disclosing the general methods of use in this invention include Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2nd ed. 1989); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and *Current Protocols in Molecular Biology* (Ausubel *et al.*, eds., 1994)).

BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 nucleic acids, polymorphic variants, orthologs, and alleles that are substantially identical to an amino acid sequence encoded by SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, or 28 can be isolated using BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 nucleic acid probes and oligonucleotides under stringent hybridization conditions, by screening libraries. Alternatively, expression libraries can be used to clone BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein, polymorphic variants, orthologs, and alleles by detecting expressed homologs immunologically with antisera or purified antibodies made against human BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 or portions thereof.

To make a cDNA library, one should choose a source that is rich in BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 RNA. The mRNA is then made into cDNA using reverse transcriptase, ligated into a recombinant vector, and transfected into a recombinant host for propagation, screening and cloning. Methods for making and screening cDNA libraries are well known (*see, e.g.*, Gubler & Hoffman, *Gene* 25:263-269 (1983); Sambrook *et al.*, *supra*; Ausubel *et al.*, *supra*).

For a genomic library, the DNA is extracted from the tissue and either mechanically sheared or enzymatically digested to yield fragments of about 12-20 kb. The fragments are then separated by gradient centrifugation from undesired sizes and are constructed in bacteriophage lambda vectors. These vectors and phage are packaged *in vitro*. Recombinant phage are analyzed by plaque hybridization as described in Benton & Davis, *Science* 196:180-182 (1977). Colony hybridization is carried out as generally described in Grunstein *et al.*, *Proc. Natl. Acad. Sci. USA.*, 72:3961-3965 (1975).

An alternative method of isolating BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 nucleic acid and its orthologs, alleles, mutants, polymorphic variants, and conservatively modified variants combines the use of synthetic oligonucleotide primers and amplification of an RNA or DNA template (*see* U.S. Patents 4,683,195 and 4,683,202; *PCR Protocols: A Guide to Methods and Applications* (Innis *et al.*, eds, 1990)). Methods such as polymerase chain reaction (PCR) and ligase chain reaction (LCR) can be used to amplify nucleic acid sequences of human BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 directly from mRNA, from cDNA, from genomic libraries or cDNA libraries. Degenerate oligonucleotides can be designed to amplify BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 homologs using the sequences provided herein. Restriction endonuclease sites can be incorporated into the primers. Polymerase chain reaction or other *in vitro* amplification methods may also be useful, for example, to clone nucleic acid sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 encoding mRNA in physiological samples, for nucleic acid sequencing, or for other purposes. Genes amplified by the PCR reaction can be purified from agarose gels and cloned into an appropriate vector.

Gene expression of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 can also be analyzed by techniques known in the art, e.g., reverse transcription and amplification of mRNA, isolation of total RNA or poly A^+

RNA, northern blotting, dot blotting, *in situ* hybridization, RNase protection, high density polynucleotide array technology, e.g., and the like.

Nucleic acids encoding BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein can be used with high density oligonucleotide array technology (e.g., GeneChip™) to identify BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein, orthologs, alleles, conservatively modified variants, and polymorphic variants in this invention. In the case where the homologs being identified are linked to modulation of cellular proliferation, they can be used with GeneChip™ as a diagnostic tool in detecting the disease in a biological sample, *see, e.g.*, Gunthand *et al., AIDS Res. Hum. Retroviruses* 14: 869-876 (1998); Kozal *et al., Nat. Med.* 2:753-759 (1996); Matson *et al., Anal. Biochem.* 224:110-106 (1995); Lockhart *et al., Nat. Biotechnol.* 14:1675-1680 (1996); Gingeras *et al., Genome Res.* 8:435-448 (1998); Hacia *et al., Nucleic Acids Res.* 26:3865-3866 (1998).

The gene for BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 is typically cloned into intermediate vectors before transformation into prokaryotic or eukaryotic cells for replication and/or expression. These intermediate vectors are typically prokaryote vectors, e.g., plasmids, or shuttle vectors.

EXPRESSION IN PROKARYOTES AND EUKARYOTES

To obtain high level expression of a cloned gene, such as those cDNAs encoding BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1, one typically subclones BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 into an expression vector that contains a strong promoter to direct transcription, a transcription/translation terminator, and if for a nucleic acid encoding a protein, a ribosome binding site for translational initiation. Suitable bacterial promoters are well known in the art and described, e.g., in Sambrook *et al.*, and Ausubel *et al, supra.* Bacterial expression systems for expressing the BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein are available in, e.g., *E.*

coli, Bacillus sp., and *Salmonella* (Palva *et al.*, *Gene* 22:229-235 (1983); Mosbach *et al.*, *Nature* 302:543-545 (1983). Kits for such expression systems are commercially available. Eukaryotic expression systems for mammalian cells, yeast, and insect cells are well known in the art and are also commercially available. In one preferred embodiment, retroviral expression systems are used in the present invention.

Selection of the promoter used to direct expression of a heterologous nucleic acid depends on the particular application. The promoter is preferably positioned about the same distance from the heterologous transcription start site as it is from the transcription start site in its natural setting. As is known in the art, however, some variation in this distance can be accommodated without loss of promoter function.

In addition to the promoter, the expression vector typically contains a transcription unit or expression cassette that contains all the additional elements required for the expression of the BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, MEMB transmembrane 4 superfamily ER 1, OR ERCC1 encoding nucleic acid in host cells. A typical expression cassette thus contains a promoter operably linked to the nucleic acid sequence encoding BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 and signals required for efficient polyadenylation of the transcript, ribosome binding sites, and translation termination. Additional elements of the cassette may include enhancers and, if genomic DNA is used as the structural gene, introns with functional splice donor and acceptor sites.

In addition to a promoter sequence, the expression cassette should also contain a transcription termination region downstream of the structural gene to provide for efficient termination. The termination region may be obtained from the same gene as the promoter sequence or may be obtained from different genes.

The particular expression vector used to transport the genetic information into the cell is not particularly critical. Any of the conventional vectors used for expression in eukaryotic or prokaryotic cells may be used. Standard bacterial expression vectors include plasmids such as pBR322 based plasmids, pSKF, pET23D, and fusion expression systems such as MBP, GST, and LacZ. Epitope tags can also be added to recombinant proteins to provide convenient methods of isolation, e.g., c-myc. Sequence tags may be included in an expression cassette for nucleic acid rescue. Markers such as fluorescent proteins, green or

red fluorescent protein, β-gal, CAT, and the like can be included in the vectors as markers for vector transduction.

Expression vectors containing regulatory elements from eukaryotic viruses are typically used in eukaryotic expression vectors, e.g., SV40 vectors, papilloma virus vectors, retroviral vectors, and vectors derived from Epstein-Barr virus. Other exemplary eukaryotic vectors include pMSG, pAV009/A$^+$, pMTO10/A$^+$, pMAMneo-5, baculovirus pDSVE, and any other vector allowing expression of proteins under the direction of the CMV promoter, SV40 early promoter, SV40 later promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedrin promoter, or other promoters shown effective for expression in eukaryotic cells.

Expression of proteins from eukaryotic vectors can be also be regulated using inducible promoters. With inducible promoters, expression levels are tied to the concentration of inducing agents, such as tetracycline or ecdysone, by the incorporation of response elements for these agents into the promoter. Generally, high level expression is obtained from inducible promoters only in the presence of the inducing agent; basal expression levels are minimal.

In one embodiment, the vectors of the invention have a regulatable promoter, e.g., tet-regulated systems and the RU-486 system (*see, e.g.*, Gossen & Bujard, *Proc. Nat'l Acad. Sci. USA* 89:5547 (1992); Oligino *et al., Gene Ther.* 5:491-496 (1998); Wang *et al., Gene Ther.* 4:432-441 (1997); Neering *et al., Blood* 88:1147-1155 (1996); and Rendahl *et al., Nat. Biotechnol.* 16:757-761 (1998)). These impart small molecule control on the expression of the candidate target nucleic acids. This beneficial feature can be used to determine that a desired phenotype is caused by a transfected cDNA rather than a somatic mutation.

Some expression systems have markers that provide gene amplification such as thymidine kinase and dihydrofolate reductase. Alternatively, high yield expression systems not involving gene amplification are also suitable, such as using a baculovirus vector in insect cells, with a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 encoding sequence under the direction of the polyhedrin promoter or other strong baculovirus promoters.

The elements that are typically included in expression vectors also include a replicon that functions in *E. coli*, a gene encoding antibiotic resistance to permit selection of bacteria that harbor recombinant plasmids, and unique restriction sites in nonessential regions of the plasmid to allow insertion of eukaryotic sequences. The particular antibiotic resistance

gene chosen is not critical, any of the many resistance genes known in the art are suitable. The prokaryotic sequences are preferably chosen such that they do not interfere with the replication of the DNA in eukaryotic cells, if necessary.

Standard transfection methods are used to produce bacterial, mammalian, yeast or insect cell lines that express large quantities of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein, which are then purified using standard techniques (*see, e.g.*, Colley *et al.*, *J. Biol. Chem.* 264:17619-17622 (1989); *Guide to Protein Purification*, in *Methods in Enzymology*, vol. 182 (Deutscher, ed., 1990)). Transformation of eukaryotic and prokaryotic cells are performed according to standard techniques (*see, e.g.*, Morrison, *J. Bact.* 132:349-351 (1977); Clark-Curtiss & Curtiss, *Methods in Enzymology* 101:347-362 (Wu *et al.*, eds, 1983).

Any of the well-known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, biolistics, liposomes, microinjection, plasma vectors, viral vectors and any of the other well known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (*see, e.g.*, Sambrook *et al., supra*). It is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one gene into the host cell capable of expressing BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1.

After the expression vector is introduced into the cells, the transfected cells are cultured under conditions favoring expression of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1, which is recovered from the culture using standard techniques identified below.

PURIFICATION OF BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, ALDEHYDE DEHYDROGENASE, PYRUVATE KINASE, G6PD, HCDR-3, DDX21, ARK2, TRANSMEMBRANE 4 SUPERFAMILY MEMBER 1, OR ERCC1 POLYPEPTIDES

Either naturally occurring or recombinant BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 can be purified for use in

functional assays. Naturally occurring BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 can be purified, e.g., from human tissue. Recombinant BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 can be purified from any suitable expression system.

The BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein may be purified to substantial purity by standard techniques, including selective precipitation with such substances as ammonium sulfate; column chromatography, immunopurification methods, and others (*see, e.g.*, Scopes, *Protein Purification: Principles and Practice* (1982); U.S. Patent No. 4,673,641; Ausubel *et al., supra;* and Sambrook *et al., supra*).

A number of procedures can be employed when recombinant BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein is being purified. For example, proteins having established molecular adhesion properties can be reversible fused to the BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein. With the appropriate ligand or substrate, e.g., antiphospho S/T antibodies or anti-BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 antibodies, BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein can be selectively adsorbed to a purification column and then freed from the column in a relatively pure form. The fused protein is then removed by enzymatic activity. Finally, BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein could be purified using immunoaffinity columns. Recombinant BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein can be purified from any suitable source, include yeast, insect, bacterial, and mammalian cells.

A. Purification of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 from recombinant bacteria

Recombinant proteins are expressed by transformed bacteria in large amounts, typically after promoter induction; but expression can be constitutive. Promoter induction with IPTG is one example of an inducible promoter system. Bacteria are grown according to standard procedures in the art. Fresh or frozen bacteria cells are used for isolation of protein.

Proteins expressed in bacteria may form insoluble aggregates ("inclusion bodies"). Several protocols are suitable for purification of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein inclusion bodies. For example, purification of inclusion bodies typically involves the extraction, separation and/or purification of inclusion bodies by disruption of bacterial cells, e.g., by incubation in a buffer of 50 mM TRIS/HCL pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 0.1 mM ATP, and 1 mM PMSF. The cell suspension can be lysed using 2-3 passages through a French Press, homogenized using a Polytron (Brinkman Instruments) or sonicated on ice. Alternate methods of lysing bacteria are apparent to those of skill in the art (*see, e.g.*, Sambrook *et al.*, *supra*; Ausubel *et al.*, *supra*).

If necessary, the inclusion bodies are solubilized, and the lysed cell suspension is typically centrifuged to remove unwanted insoluble matter. Proteins that formed the inclusion bodies may be renatured by dilution or dialysis with a compatible buffer. Suitable solvents include, but are not limited to urea (from about 4 M to about 8 M), formamide (at least about 80%, volume/volume basis), and guanidine hydrochloride (from about 4 M to about 8 M). Some solvents which are capable of solubilizing aggregate-forming proteins, for example SDS (sodium dodecyl sulfate), 70% formic acid, are inappropriate for use in this procedure due to the possibility of irreversible denaturation of the proteins, accompanied by a lack of immunogenicity and/or activity. Although guanidine hydrochloride and similar agents are denaturants, this denaturation is not irreversible and renaturation may occur upon removal (by dialysis, for example) or dilution of the denaturant, allowing re-formation of immunologically and/or biologically active protein. Other suitable buffers are known to those skilled in the art. Human BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 proteins are separated from other bacterial proteins by standard separation techniques, e.g., with Ni-NTA agarose resin.

Alternatively, it is possible to purify BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein from bacteria periplasm. After lysis of the bacteria, when the BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein exported into the periplasm of the bacteria, the periplasmic fraction of the bacteria can be isolated by cold osmotic shock in addition to other methods known to skill in the art. To isolate recombinant proteins from the periplasm, the bacterial cells are centrifuged to form a pellet. The pellet is resuspended in a buffer containing 20% sucrose. To lyse the cells, the bacteria are centrifuged and the pellet is resuspended in ice-cold 5 mM $MgSO_4$ and kept in an ice bath for approximately 10 minutes. The cell suspension is centrifuged and the supernatant decanted and saved. The recombinant proteins present in the supernatant can be separated from the host proteins by standard separation techniques well known to those of skill in the art.

B. Standard protein separation techniques for purifying BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 proteins

Solubility fractionation

Often as an initial step, particularly if the protein mixture is complex, an initial salt fractionation can separate many of the unwanted host cell proteins (or proteins derived from the cell culture media) from the recombinant protein of interest. The preferred salt is ammonium sulfate. Ammonium sulfate precipitates proteins by effectively reducing the amount of water in the protein mixture. Proteins then precipitate on the basis of their solubility. The more hydrophobic a protein is, the more likely it is to precipitate at lower ammonium sulfate concentrations. A typical protocol includes adding saturated ammonium sulfate to a protein solution so that the resultant ammonium sulfate concentration is between 20-30%. This concentration will precipitate the most hydrophobic of proteins. The precipitate is then discarded (unless the protein of interest is hydrophobic) and ammonium sulfate is added to the supernatant to a concentration known to precipitate the protein of interest. The precipitate is then solubilized in buffer and the excess salt removed if necessary, either through dialysis or diafiltration. Other methods that rely on solubility of proteins, such as cold ethanol precipitation, are well known to those of skill in the art and can be used to fractionate complex protein mixtures.

Size differential filtration

The molecular weight of the BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 proteins can be used to isolate it from proteins of greater and lesser size using ultrafiltration through membranes of different pore size (for example, Amicon or Millipore membranes). As a first step, the protein mixture is ultrafiltered through a membrane with a pore size that has a lower molecular weight cut-off than the molecular weight of the protein of interest. The retentate of the ultrafiltration is then ultrafiltered against a membrane with a molecular cut off greater than the molecular weight of the protein of interest. The recombinant protein will pass through the membrane into the filtrate. The filtrate can then be chromatographed as described below.

Column chromatography

The BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 proteins can also be separated from other proteins on the basis of its size, net surface charge, hydrophobicity, and affinity for ligands. In addition, antibodies raised against proteins can be conjugated to column matrices and the proteins immunopurified. All of these methods are well known in the art. It will be apparent to one of skill that chromatographic techniques can be performed at any scale and using equipment from many different manufacturers (e.g., Pharmacia Biotech).

ASSAYS FOR MODULATORS OF BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, ALDEHYDE DEHYDROGENASE, PYRUVATE KINASE, G6PD, HCDR-3, DDX21, ARK2, TRANSMEMBRANE 4 SUPERFAMILY MEMBER 1, OR ERCC1 PROTEIN

A. Assays

Modulation of a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein, and corresponding modulation of cellular, e.g., tumor cell, proliferation, can be assessed using a variety of *in vitro* and *in vivo* assays, including cell-based models. Such assays can be used to test for inhibitors and activators of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1

protein, and, consequently, inhibitors and activators of cellular proliferation, including modulators of chemotherapeutic sensitivity and toxicity. Such modulators of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein are useful for treating disorders related to pathological cell proliferation, e.g., cancer. Modulators of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein are tested using either recombinant or naturally occurring BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1, preferably human BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1.

Preferably, the BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein will have the sequence as encoded by SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 or a conservatively modified variant thereof. Alternatively, the BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein of the assay will be derived from a eukaryote and include an amino acid subsequence having substantial amino acid sequence identity to SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, or 28. Generally, the amino acid sequence identity will be at least 60%, preferably at least 65%, 70%, 75%, 80%, 85%, or 90%, most preferably at least 95%.

Measurement of cellular proliferation modulation with BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein or a cell expressing BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein, either recombinant or naturally occurring, can be performed using a variety of assays, *in vitro*, *in vivo*, and *ex vivo*, as described herein. A suitable physical, chemical or phenotypic change that affects activity, e.g., enzymatic activity such as kinase activity, cell proliferation, or ligand binding can be used to assess the influence of a test compound on the polypeptide of this invention. When the functional effects are determined

using intact cells or animals, one can also measure a variety of effects, such as, ligand binding, kinase activity, transcriptional changes to both known and uncharacterized genetic markers (e.g., northern blots), changes in cell metabolism, changes related to cellular proliferation, cell surface marker expression, DNA synthesis, marker and dye dilution assays (e.g., GFP and cell tracker assays), contact inhibition, tumor growth in nude mice, etc.

In vitro assays

Assays to identify compounds with BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 modulating activity can be performed *in vitro*. Such assays can used full length BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein or a variant thereof (*see, e.g.*, SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28), or a mutant thereof, or a fragment of a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein, such as a kinase domain. Purified recombinant or naturally occurring BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein can be used in the *in vitro* methods of the invention. In addition to purified BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein, the recombinant or naturally occurring BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein can be part of a cellular lysate or a cell membrane. As described below, the binding assay can be either solid state or soluble. Preferably, the protein or membrane is bound to a solid support, either covalently or non-covalently. Often, the *in vitro* assays of the invention are substrate or ligand binding or affinity assays, either non-competitive or competitive. Other *in vitro* assays include measuring changes in spectroscopic (e.g., fluorescence, absorbance, refractive index), hydrodynamic (e.g., shape), chromatographic, or solubility properties for the protein. Other in vitro assays include enzymatic activity assays, such as phosphorylation or autophosphorylation assays.

In one embodiment, a high throughput binding assay is performed in which the BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate

kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein or a fragment thereof is contacted with a potential modulator and incubated for a suitable amount of time. In one embodiment, the potential modulator is bound to a solid support, and the BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein is added. In another embodiment, the BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein is bound to a solid support. A wide variety of modulators can be used, as described below, including small organic molecules, peptides, antibodies, and BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 ligand analogs. A wide variety of assays can be used to identify BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1-modulator binding, including labeled protein-protein binding assays, electrophoretic mobility shifts, immunoassays, enzymatic assays such as kinase assays, and the like. In some cases, the binding of the candidate modulator is determined through the use of competitive binding assays, where interference with binding of a known ligand or substrate is measured in the presence of a potential modulator. Either the modulator or the known ligand or substrate is bound first, and then the competitor is added. After the BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein is washed, interference with binding, either of the potential modulator or of the known ligand or substrate, is determined. Often, either the potential modulator or the known ligand or substrate is labeled.

Cell-based *in vivo* assays

In another embodiment, BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein is expressed in a cell, and functional, e.g., physical and chemical or phenotypic, changes are assayed to identify BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 and modulators of cellular proliferation, e.g., tumor cell proliferation. Cells expressing BAP-1, NP95, FANCA,

DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 proteins can also be used in binding assays and enzymatic assays. Any suitable functional effect can be measured, as described herein. For example, cellular morphology (e.g., cell volume, nuclear volume, cell perimeter, and nuclear perimeter), ligand binding, kinase activity, apoptosis, cell surface marker expression, cellular proliferation, GFP positivity and dye dilution assays (e.g., cell tracker assays with dyes that bind to cell membranes), DNA synthesis assays (e.g., ^{3}H-thymidine and fluorescent DNA-binding dyes such as BrdU or Hoescht dye with FACS analysis), are all suitable assays to identify potential modulators using a cell based system. Suitable cells for such cell based assays include both primary cancer or tumor cells and cell lines, as described herein, e.g., A549 (lung), MCF7 (breast, p53 wild-type), H1299 (lung, p53 null), Hela (cervical), PC3 (prostate, p53 mutant), MDA-MB-231 (breast, p53 wild-type). Cancer cell lines can be p53 mutant, p53 null, or express wild type p53. The BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein can be naturally occurring or recombinant. Also, fragments of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 or chimeric BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 proteins with enzymatic activity can be used in cell based assays.

Cellular BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 polypeptide levels can be determined by measuring the level of protein or mRNA. The level of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, OR ERCC1 protein or proteins related to BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 are measured using immunoassays such as western blotting, ELISA and the like with an antibody that selectively binds to the BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 polypeptide or a fragment thereof. For measurement of mRNA, amplification, e.g., using PCR, LCR, or hybridization assays, e.g., northern hybridization, RNAse

protection, dot blotting, are preferred. The level of protein or mRNA is detected using directly or indirectly labeled detection agents, e.g., fluorescently or radioactively labeled nucleic acids, radioactively or enzymatically labeled antibodies, and the like, as described herein.

Alternatively, BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 expression can be measured using a reporter gene system. Such a system can be devised using a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein promoter operably linked to a reporter gene such as chloramphenicol acetyltransferase, firefly luciferase, bacterial luciferase, β-galactosidase and alkaline phosphatase. Furthermore, the protein of interest can be used as an indirect reporter via attachment to a second reporter such as red or green fluorescent protein (*see, e.g.*, Mistili & Spector, *Nature Biotechnology* 15:961-964 (1997)). The reporter construct is typically transfected into a cell. After treatment with a potential modulator, the amount of reporter gene transcription, translation, or activity is measured according to standard techniques known to those of skill in the art.

Animal models

Animal models of cellular proliferation also find use in screening for modulators of cellular proliferation. Similarly, transgenic animal technology including gene knockout technology, for example as a result of homologous recombination with an appropriate gene targeting vector, or gene overexpression, will result in the absence or increased expression of the BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein. The same technology can also be applied to make knock-out cells. When desired, tissue-specific expression or knockout of the BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein may be necessary. Transgenic animals generated by such methods find use as animal models of cellular proliferation and are additionally useful in screening for modulators of cellular proliferation.

Knock-out cells and transgenic mice can be made by insertion of a marker gene or other heterologous gene into an endogenous BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 gene site in the mouse genome via homologous recombination. Such mice can also be made by substituting an endogenous BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 with a mutated version of the BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 gene, or by mutating an endogenous BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1, e.g., by exposure to carcinogens.

A DNA construct is introduced into the nuclei of embryonic stem cells. Cells containing the newly engineered genetic lesion are injected into a host mouse embryo, which is re-implanted into a recipient female. Some of these embryos develop into chimeric mice that possess germ cells partially derived from the mutant cell line. Therefore, by breeding the chimeric mice it is possible to obtain a new line of mice containing the introduced genetic lesion (*see, e.g.,* Capecchi *et al., Science* 244:1288 (1989)). Chimeric targeted mice can be derived according to Hogan *et al.*, *Manipulating the Mouse Embryo*: *A Laboratory Manual*, Cold Spring Harbor Laboratory (1988) and *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson, ed., IRL Press, Washington, D.C., (1987).

Exemplary assays

Enzymatic activity assays-- *in vitro* or cell based

In one embodiment, enzymatic assays using BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 can be used to identify modulators of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 kinase activity, or to identify proteins that bind to BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1, e.g., BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3,

DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 substrates. Full length wild type BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1, mutant BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1, or the BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 enzymatic domain can be used in these assays. Such assays can be performed *in vitro*, using recombinant BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 or cellular lysates comprising endogenous or recombinant BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1, or can be cell-based.

Soft agar growth or colony formation in suspension

Normal cells require a solid substrate to attach and grow. When the cells are transformed, they lose this phenotype and grow detached from the substrate. For example, transformed cells can grow in stirred suspension culture or suspended in semi-solid media, such as semi-solid or soft agar. The transformed cells, when transfected with tumor suppressor genes, regenerate normal phenotype and require a solid substrate to attach and grow.

Soft agar growth or colony formation in suspension assays can be used to identify BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 modulators. Typically, transformed host cells (e.g., cells that grow on soft agar) are used in this assay. For example, RKO or HCT116 cell lines can be used. Techniques for soft agar growth or colony formation in suspension assays are described in Freshney, *Culture of Animal Cells a Manual of Basic Technique*, 3rd ed., Wiley-Liss, New York (1994), herein incorporated by reference. *See also*, the methods section of Garkavtsev *et al.* (1996), *supra*, herein incorporated by reference.

Contact inhibition and density limitation of growth

Normal cells typically grow in a flat and organized pattern in a petri dish until they touch other cells. When the cells touch one another, they are contact inhibited and stop growing. When cells are transformed, however, the cells are not contact inhibited and continue to grow to high densities in disorganized foci. Thus, the transformed cells grow to a higher saturation density than normal cells. This can be detected morphologically by the formation of a disoriented monolayer of cells or rounded cells in foci within the regular pattern of normal surrounding cells. Alternatively, labeling index with [^{3}H]-thymidine at saturation density can be used to measure density limitation of growth. *See* Freshney (1994), *supra*. The transformed cells, when contacted with cellular proliferation modulators, regenerate a normal phenotype and become contact inhibited and would grow to a lower density.

Contact inhibition and density limitation of growth assays can be used to identify BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 modulators which are capable of inhibiting abnormal proliferation and transformation in host cells. Typically, transformed host cells (e.g., cells that are not contact inhibited) are used in this assay. For example, RKO or HCT116 cell lines can be used. In this assay, labeling index with [^{3}H]-thymidine at saturation density is a preferred method of measuring density limitation of growth. Transformed host cells are contacted with a potential BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 modulator and are grown for 24 hours at saturation density in non-limiting medium conditions. The percentage of cells labeling with [^{3}H]-thymidine is determined autoradiographically. *See*, Freshney (1994), *supra*. The host cells contacted with a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 modulator would give arise to a lower labeling index compared to control (e.g., transformed host cells transfected with a vector lacking an insert).

Growth factor or serum dependence

Growth factor or serum dependence can be used as an assay to identify BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1

modulators. Transformed cells have a lower serum dependence than their normal counterparts (*see, e.g.*, Temin, *J. Natl. Cancer Insti.* 37:167-175 (1966); Eagle *et al.*, *J. Exp. Med.* 131:836-879 (1970)); Freshney, *supra*. This is in part due to release of various growth factors by the transformed cells. When transformed cells are contacted with a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 modulator, the cells would reacquire serum dependence and would release growth factors at a lower level.

Tumor specific markers levels

Tumor cells release an increased amount of certain factors (hereinafter "tumor specific markers") than their normal counterparts. For example, plasminogen activator (PA) is released from human glioma at a higher level than from normal brain cells (*see, e.g.*, Gullino, *Angiogenesis, tumor vascularization, and potential interference with tumor growth.* In Mihich (ed.): "Biological Responses in Cancer." New York, Academic Press, pp. 178-184 (1985)). Similarly, tumor angiogenesis factor (TAF) is released at a higher level in tumor cells than their normal counterparts. *See, e.g.*, Folkman, Angiogenesis and cancer, *Sem Cancer Biol.* (1992)).

Tumor specific markers can be assayed to identify BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 modulators which decrease the level of release of these markers from host cells. Typically, transformed or tumorigenic host cells are used. Various techniques which measure the release of these factors are described in Freshney (1994), *supra*. Also, *see*, Unkless *et al.* , *J. Biol. Chem.* 249:4295-4305 (1974); Strickland & Beers, *J. Biol. Chem.* 251:5694-5702 (1976); Whur *et al.*, *Br. J. Cancer* 42:305-312 (1980); Gulino, *Angiogenesis, tumor vascularization, and potential interference with tumor growth.* In Mihich, E. (ed): "Biological Responses in Cancer." New York, Plenum (1985); Freshney *Anticancer Res.* 5:111-130 (1985).

Invasiveness into Matrigel

The degree of invasiveness into Matrigel or some other extracellular matrix constituent can be used as an assay to identify BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 modulators which are capable of

inhibiting abnormal cell proliferation and tumor growth. Tumor cells exhibit a good correlation between malignancy and invasiveness of cells into Matrigel or some other extracellular matrix constituent. In this assay, tumorigenic cells are typically used as host cells. Therefore, BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 modulators can be identified by measuring changes in the level of invasiveness between the host cells before and after the introduction of potential modulators. If a compound modulates BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1, its expression in tumorigenic host cells would affect invasiveness.

Techniques described in Freshney (1994), *supra*, can be used. Briefly, the level of invasion of host cells can be measured by using filters coated with Matrigel or some other extracellular matrix constituent. Penetration into the gel, or through to the distal side of the filter, is rated as invasiveness, and rated histologically by number of cells and distance moved, or by prelabeling the cells with ^{125}I and counting the radioactivity on the distal side of the filter or bottom of the dish. *See, e.g.*, Freshney (1984), *supra*.

Apoptosis analysis

Apoptosis analysis can be used as an assay to identify BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 modulators. In this assay, cell lines, such as RKO or HCT116, can be used to screen BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 modulators. Cells are contacted with a putative BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 modulator. The cells can be co-transfected with a construct comprising a marker gene, such as a gene that encodes green fluorescent protein, or a cell tracker dye. The apoptotic change can be determined using methods known in the art, such as DAPI staining and TUNEL assay using fluorescent microscope. For TUNEL assay, commercially available kit can be used (e.g., Fluorescein FragEL DNA Fragmentation Detection Kit (Oncogene Research Products, Cat.# QIA39) + Tetramethyl-rhodamine-5-dUTP (Roche, Cat. # 1534 378)). Cells contacted with BAP-1, NP95, FANCA, DDX9,

IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 modulators would exhibit, e.g., an increased apoptosis compared to control.

G_0/G_1 cell cycle arrest analysis

G_0/G_1 cell cycle arrest can be used as an assay to identify BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 modulators. In this assay, cell lines, such as RKO or HCT116, can be used to screen BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 modulators. The cells can be co-transfected with a construct comprising a marker gene, such as a gene that encodes green fluorescent protein, or a cell tracker dye. Methods known in the art can be used to measure the degree of G_1 cell cycle arrest. For example, a propidium iodide signal can be used as a measure for DNA content to determine cell cycle profiles on a flow cytometer. The percent of the cells in each cell cycle can be calculated. Cells contacted with a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 modulator would exhibit, e.g., a higher number of cells that are arrested in G_0/G_1 phase compared to control.

Tumor growth *in vivo*

Effects of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 modulators on cell growth can be tested in transgenic or immune-suppressed mice. Knock-out transgenic mice can be made, in which the endogenous BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 gene is disrupted. Such knock-out mice can be used to study effects of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1, e.g., as a cancer model, as a means of assaying *in vivo* for compounds that modulate BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1, and to test the effects of

restoring a wild-type or mutant BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 to a knock-out mice.

Knock-out cells and transgenic mice can be made by insertion of a marker gene or other heterologous gene into the endogenous BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 gene site in the mouse genome via homologous recombination. Such mice can also be made by substituting the endogenous BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 with a mutated version of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1, or by mutating the endogenous BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1, e.g., by exposure to carcinogens.

A DNA construct is introduced into the nuclei of embryonic stem cells. Cells containing the newly engineered genetic lesion are injected into a host mouse embryo, which is re-implanted into a recipient female. Some of these embryos develop into chimeric mice that possess germ cells partially derived from the mutant cell line. Therefore, by breeding the chimeric mice it is possible to obtain a new line of mice containing the introduced genetic lesion (*see, e.g.,* Capecchi *et al., Science* 244:1288 (1989)). Chimeric targeted mice can be derived according to Hogan *et al., Manipulating the Mouse Embryo: A Laboratory Manual*, Cold Spring Harbor Laboratory (1988) and *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson, ed., IRL Press, Washington, D.C., (1987). These knock-out mice can be used as hosts to test the effects of various BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 modulators on cell growth.

Alternatively, various immune-suppressed or immune-deficient host animals can be used. For example, genetically athymic "nude" mouse (*see, e.g.*, Giovanella *et al., J. Natl. Cancer Inst.* 52:921 (1974)), a SCID mouse, a thymectomized mouse, or an irradiated mouse (*see, e.g.*, Bradley *et al., Br. J. Cancer* 38:263 (1978); Selby *et al., Br. J. Cancer* 41:52 (1980)) can be used as a host. Transplantable tumor cells (typically about 10^6 cells) injected into isogenic hosts will produce invasive tumors in a high proportions of cases, while

normal cells of similar origin will not. Hosts are treated with BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 modulators, e.g., by injection. After a suitable length of time, preferably 4-8 weeks, tumor growth is measured (e.g., by volume or by its two largest dimensions) and compared to the control. Tumors that have statistically significant reduction (using, e.g., Student's T test) are said to have inhibited growth. Using reduction of tumor size as an assay, BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 modulators which are capable, e.g., of inhibiting abnormal cell proliferation can be identified.

B. Modulators

The compounds tested as modulators of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein can be any small organic molecule, or a biological entity, such as a protein, e.g., an antibody or peptide, a sugar, a nucleic acid, e.g., an antisense oligonucleotide or a ribozyme, or a lipid. Alternatively, modulators can be genetically altered versions of a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein. Typically, test compounds will be small organic molecules, peptides, circular peptides, RNAi, antisense molecules, ribozymes, and lipids.

Essentially any chemical compound can be used as a potential modulator or ligand in the assays of the invention, although most often compounds can be dissolved in aqueous or organic (especially DMSO-based) solutions are used. The assays are designed to screen large chemical libraries by automating the assay steps and providing compounds from any convenient source to assays, which are typically run in parallel (e.g., in microtiter formats on microtiter plates in robotic assays). It will be appreciated that there are many suppliers of chemical compounds, including Sigma (St. Louis, MO), Aldrich (St. Louis, MO), Sigma-Aldrich (St. Louis, MO), Fluka Chemika-Biochemica Analytika (Buchs Switzerland) and the like.

In one preferred embodiment, high throughput screening methods involve providing a combinatorial small organic molecule or peptide library containing a large number of potential therapeutic compounds (potential modulator or ligand compounds).

Such "combinatorial chemical libraries" or "ligand libraries" are then screened in one or more assays, as described herein, to identify those library members (particular chemical species or subclasses) that display a desired characteristic activity. The compounds thus identified can serve as conventional "lead compounds" or can themselves be used as potential or actual therapeutics.

A combinatorial chemical library is a collection of diverse chemical compounds generated by either chemical synthesis or biological synthesis, by combining a number of chemical "building blocks" such as reagents. For example, a linear combinatorial chemical library such as a polypeptide library is formed by combining a set of chemical building blocks (amino acids) in every possible way for a given compound length (i.e., the number of amino acids in a polypeptide compound). Millions of chemical compounds can be synthesized through such combinatorial mixing of chemical building blocks.

Preparation and screening of combinatorial chemical libraries is well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (*see, e.g.,* U.S. Patent 5,010,175, Furka, *Int. J. Pept. Prot. Res.* 37:487-493 (1991) and Houghton *et al.*, *Nature* 354:84-88 (1991)). Other chemistries for generating chemical diversity libraries can also be used. Such chemistries include, but are not limited to: peptoids (e.g., PCT Publication No. WO 91/19735), encoded peptides (e.g., PCT Publication No. WO 93/20242), random bio-oligomers (e.g., PCT Publication No. WO 92/00091), benzodiazepines (e.g., U.S. Pat. No. 5,288,514), diversomers such as hydantoins, benzodiazepines and dipeptides (Hobbs *et al.*, *Proc. Nat. Acad. Sci. USA* 90:6909-6913 (1993)), vinylogous polypeptides (Hagihara *et al.*, *J. Amer. Chem. Soc.* 114:6568 (1992)), nonpeptidal peptidomimetics with glucose scaffolding (Hirschmann *et al.*, *J. Amer. Chem. Soc.* 114:9217-9218 (1992)), analogous organic syntheses of small compound libraries (Chen *et al.*, *J. Amer. Chem. Soc.* 116:2661 (1994)), oligocarbamates (Cho *et al.*, *Science* 261:1303 (1993)), and/or peptidyl phosphonates (Campbell *et al.*, *J. Org. Chem.* 59:658 (1994)), nucleic acid libraries (*see* Ausubel, Berger and Sambrook, all *supra*), peptide nucleic acid libraries (*see, e.g.,* U.S. Patent 5,539,083), antibody libraries (*see, e.g.,* Vaughn *et al.*, *Nature Biotechnology*, 14(3):309-314 (1996) and PCT/US96/10287), carbohydrate libraries (*see, e.g.,* Liang *et al.*, *Science*, 274:1520-1522 (1996) and U.S. Patent 5,593,853), small organic molecule libraries (*see, e.g.,* benzodiazepines, Baum C&EN, Jan 18, page 33 (1993); isoprenoids, U.S. Patent 5,569,588; thiazolidinones and metathiazanones, U.S. Patent 5,549,974; pyrrolidines, U.S. Patents 5,525,735 and 5,519,134; morpholino compounds, U.S. Patent 5,506,337; benzodiazepines, 5,288,514, and the like).

Devices for the preparation of combinatorial libraries are commercially available (*see, e.g.*, 357 MPS, 390 MPS, Advanced Chem Tech, Louisville KY, Symphony, Rainin, Woburn, MA, 433A Applied Biosystems, Foster City, CA, 9050 Plus, Millipore, Bedford, MA). In addition, numerous combinatorial libraries are themselves commercially available (*see, e.g.*, ComGenex, Princeton, N.J., Asinex, Moscow, Ru, Tripos, Inc., St. Louis, MO, ChemStar, Ltd, Moscow, RU, 3D Pharmaceuticals, Exton, PA, Martek Biosciences, Columbia, MD, etc.).

C. Solid state and soluble high throughput assays

In one embodiment the invention provides soluble assays using a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein, or a cell or tissue expressing a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein, either naturally occurring or recombinant. In another embodiment, the invention provides solid phase based *in vitro* assays in a high throughput format, where the BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein or BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 substrate is attached to a solid phase. Any one of the assays described herein can be adapted for high throughput screening.

In the high throughput assays of the invention, either soluble or solid state, it is possible to screen up to several thousand different modulators or ligands in a single day. This methodology can be used for BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 proteins *in vitro*, or for cell-based or membrane-based assays comprising a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein. In particular, each well of a microtiter plate can be used to run a separate assay against a selected potential modulator, or, if concentration or incubation time effects are to be observed, every 5-10 wells can test a single modulator. Thus, a single standard microtiter plate can assay about 100 (e.g., 96) modulators. If 1536 well plates are used, then a single plate can easily assay from about 100- about 1500 different

compounds. It is possible to assay many plates per day; assay screens for up to about 6,000, 20,000, 50,000, or more than 100,000 different compounds are possible using the integrated systems of the invention.

For a solid state reaction, the protein of interest or a fragment thereof, e.g., an extracellular domain, or a cell or membrane comprising the protein of interest or a fragment thereof as part of a fusion protein can be bound to the solid state component, directly or indirectly, via covalent or non covalent linkage. A tag for covalent or non-covalent binding can be any of a variety of components. In general, a molecule which binds the tag (a tag binder) is fixed to a solid support, and the tagged molecule of interest is attached to the solid support by interaction of the tag and the tag binder.

A number of tags and tag binders can be used, based upon known molecular interactions well described in the literature. For example, where a tag has a natural binder, for example, biotin, protein A, or protein G, it can be used in conjunction with appropriate tag binders (avidin, streptavidin, neutravidin, the Fc region of an immunoglobulin, etc.) Antibodies to molecules with natural binders such as biotin are also widely available and appropriate tag binders; *see,* SIGMA Immunochemicals 1998 catalogue SIGMA, St. Louis MO).

Similarly, any haptenic or antigenic compound can be used in combination with an appropriate antibody to form a tag/tag binder pair. Thousands of specific antibodies are commercially available and many additional antibodies are described in the literature. For example, in one common configuration, the tag is a first antibody and the tag binder is a second antibody which recognizes the first antibody. In addition to antibody-antigen interactions, receptor-ligand interactions are also appropriate as tag and tag-binder pairs. For example, agonists and antagonists of cell membrane receptors (e.g., cell receptor-ligand interactions such as transferrin, c-kit, viral receptor ligands, cytokine receptors, chemokine receptors, interleukin receptors, immunoglobulin receptors and antibodies, the cadherein family, the integrin family, the selectin family, and the like; *see, e.g.,* Pigott & Power, *The Adhesion Molecule Facts Book I* (1993). Similarly, toxins and venoms, viral epitopes, hormones (e.g., opiates, steroids, etc.), intracellular receptors (e.g. which mediate the effects of various small ligands, including steroids, thyroid hormone, retinoids and vitamin D; peptides), drugs, lectins, sugars, nucleic acids (both linear and cyclic polymer configurations), oligosaccharides, proteins, phospholipids and antibodies can all interact with various cell receptors.

Synthetic polymers, such as polyurethanes, polyesters, polycarbonates, polyureas, polyamides, polyethyleneimines, polyarylene sulfides, polysiloxanes, polyimides, and polyacetates can also form an appropriate tag or tag binder. Many other tag/tag binder pairs are also useful in assay systems described herein, as would be apparent to one of skill upon review of this disclosure.

Common linkers such as peptides, polyethers, and the like can also serve as tags, and include polypeptide sequences, such as poly gly sequences of between about 5 and 200 amino acids. Such flexible linkers are known to persons of skill in the art. For example, poly(ethelyne glycol) linkers are available from Shearwater Polymers, Inc. Huntsville, Alabama. These linkers optionally have amide linkages, sulfhydryl linkages, or heterofunctional linkages.

Tag binders are fixed to solid substrates using any of a variety of methods currently available. Solid substrates are commonly derivatized or functionalized by exposing all or a portion of the substrate to a chemical reagent which fixes a chemical group to the surface which is reactive with a portion of the tag binder. For example, groups which are suitable for attachment to a longer chain portion would include amines, hydroxyl, thiol, and carboxyl groups. Aminoalkylsilanes and hydroxyalkylsilanes can be used to functionalize a variety of surfaces, such as glass surfaces. The construction of such solid phase biopolymer arrays is well described in the literature. *See, e.g.*, Merrifield, *J. Am. Chem. Soc.* 85:2149-2154 (1963) (describing solid phase synthesis of, e.g., peptides); Geysen *et al., J. Immun. Meth.* 102:259-274 (1987) (describing synthesis of solid phase components on pins); Frank & Doring, *Tetrahedron* 44:60316040 (1988) (describing synthesis of various peptide sequences on cellulose disks); Fodor *et al., Science*, 251:767-777 (1991); Sheldon *et al., Clinical Chemistry* 39(4):718-719 (1993); and Kozal *et al., Nature Medicine* 2(7):753759 (1996) (all describing arrays of biopolymers fixed to solid substrates). Non-chemical approaches for fixing tag binders to substrates include other common methods, such as heat, cross-linking by UV radiation, and the like.

IMMUNOLOGICAL DETECTION OF BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, ALDEHYDE DEHYDROGENASE, PYRUVATE KINASE, G6PD, HCDR-3, DDX21, ARK2, TRANSMEMBRANE 4 SUPERFAMILY MEMBER 1, OR ERCC1 POLYPEPTIDES

In addition to the detection of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2,

transmembrane 4 superfamily member 1, or ERCC1 gene and gene expression using nucleic acid hybridization technology, one can also use immunoassays to detect BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 proteins of the invention. Such assays are useful for screening for modulators of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1, as well as for therapeutic and diagnostic applications. Immunoassays can be used to qualitatively or quantitatively analyze BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein. A general overview of the applicable technology can be found in Harlow & Lane, *Antibodies: A Laboratory Manual* (1988).

A. Production of antibodies

Methods of producing polyclonal and monoclonal antibodies that react specifically with the BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 proteins are known to those of skill in the art (*see, e.g.*, Coligan, *Current Protocols in Immunology* (1991); Harlow & Lane, *supra*; Goding, *Monoclonal Antibodies: Principles and Practice* (2d ed. 1986); and Kohler & Milstein, *Nature* 256:495-497 (1975). Such techniques include antibody preparation by selection of antibodies from libraries of recombinant antibodies in phage or similar vectors, as well as preparation of polyclonal and monoclonal antibodies by immunizing rabbits or mice (*see, e.g.*, Huse *et al., Science* 246:1275-1281 (1989); Ward *et al., Nature* 341:544-546 (1989)).

A number of immunogens comprising portions of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein may be used to produce antibodies specifically reactive with BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, Aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein. For example, recombinant BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein or an antigenic fragment thereof, can be isolated as described herein. Recombinant protein can be expressed in eukaryotic or prokaryotic cells as described above, and purified

as generally described above. Recombinant protein is the preferred immunogen for the production of monoclonal or polyclonal antibodies. Alternatively, a synthetic peptide derived from the sequences disclosed herein and conjugated to a carrier protein can be used an immunogen. Naturally occurring protein may also be used either in pure or impure form. The product is then injected into an animal capable of producing antibodies. Either monoclonal or polyclonal antibodies may be generated, for subsequent use in immunoassays to measure the protein.

Methods of production of polyclonal antibodies are known to those of skill in the art. An inbred strain of mice (e.g., BALB/C mice) or rabbits is immunized with the protein using a standard adjuvant, such as Freund's adjuvant, and a standard immunization protocol. The animal's immune response to the immunogen preparation is monitored by taking test bleeds and determining the titer of reactivity to the beta subunits. When appropriately high titers of antibody to the immunogen are obtained, blood is collected from the animal and antisera are prepared. Further fractionation of the antisera to enrich for antibodies reactive to the protein can be done if desired (*see,* Harlow & Lane, *supra*).

Monoclonal antibodies may be obtained by various techniques familiar to those skilled in the art. Briefly, spleen cells from an animal immunized with a desired antigen are immortalized, commonly by fusion with a myeloma cell (*see,* Kohler & Milstein, *Eur. J. Immunol.* 6:511-519 (1976)). Alternative methods of immortalization include transformation with Epstein Barr Virus, oncogenes, or retroviruses, or other methods well known in the art. Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies produced by such cells may be enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host. Alternatively, one may isolate DNA sequences which encode a monoclonal antibody or a binding fragment thereof by screening a DNA library from human B cells according to the general protocol outlined by Huse, *et al.*, *Science* 246:1275-1281 (1989).

Monoclonal antibodies and polyclonal sera are collected and titered against the immunogen protein in an immunoassay, for example, a solid phase immunoassay with the immunogen immobilized on a solid support. Typically, polyclonal antisera with a titer of 10^4 or greater are selected and tested for their cross reactivity against non- BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 proteins, using a competitive binding immunoassay. Specific polyclonal antisera and monoclonal

antibodies will usually bind with a K_d of at least about 0.1 mM, more usually at least about 1 μM, preferably at least about 0.1 μM or better, and most preferably, 0.01 μM or better. Antibodies specific only for a particular BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 ortholog, such as human BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1, can also be made, by subtracting out other cross-reacting orthologs from a species such as a non-human mammal. In this manner, antibodies that bind only to BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein may be obtained.

Once the specific antibodies against BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein are available, the protein can be detected by a variety of immunoassay methods. In addition, the antibody can be used therapeutically as a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 modulators. For a review of immunological and immunoassay procedures, see *Basic and Clinical Immunology* (Stites & Terr eds., 7th ed. 1991). Moreover, the immunoassays of the present invention can be performed in any of several configurations, which are reviewed extensively in Enzyme Immunoassay (Maggio, ed., 1980); and Harlow & Lane, *supra*.

B. Immunological binding assays

BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein can be detected and/or quantified using any of a number of well recognized immunological binding assays (*see, e.g.,* U.S. Patents 4,366,241; 4,376,110; 4,517,288; and 4,837,168). For a review of the general immunoassays, see also *Methods in Cell Biology: Antibodies in Cell Biology*, volume 37 (Asai, ed. 1993); *Basic and Clinical Immunology* (Stites & Terr, eds., 7th ed. 1991). Immunological binding assays (or immunoassays) typically use an antibody that specifically binds to a protein or antigen of choice (in this case the BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4

superfamily member 1, or ERCC1 protein or antigenic subsequence thereof). The antibody (e.g., anti- BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1) may be produced by any of a number of means well known to those of skill in the art and as described above.

Immunoassays also often use a labeling agent to specifically bind to and label the complex formed by the antibody and antigen. The labeling agent may itself be one of the moieties comprising the antibody/antigen complex. Thus, the labeling agent may be a labeled BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 or a labeled anti- BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 antibody. Alternatively, the labeling agent may be a third moiety, such a secondary antibody, that specifically binds to the antibody/ BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 complex (a secondary antibody is typically specific to antibodies of the species from which the first antibody is derived). Other proteins capable of specifically binding immunoglobulin constant regions, such as protein A or protein G may also be used as the label agent. These proteins exhibit a strong non-immunogenic reactivity with immunoglobulin constant regions from a variety of species (*see, e.g.,* Kronval *et al., J. Immunol.* 111:1401-1406 (1973); Akerstrom *et al., J. Immunol.* 135:2589-2542 (1985)). The labeling agent can be modified with a detectable moiety, such as biotin, to which another molecule can specifically bind, such as streptavidin. A variety of detectable moieties are well known to those skilled in the art.

Throughout the assays, incubation and/or washing steps may be required after each combination of reagents. Incubation steps can vary from about 5 seconds to several hours, optionally from about 5 minutes to about 24 hours. However, the incubation time will depend upon the assay format, antigen, volume of solution, concentrations, and the like. Usually, the assays will be carried out at ambient temperature, although they can be conducted over a range of temperatures, such as 10°C to 40°C.

Non-competitive assay formats

Immunoassays for detecting BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 in samples may be either competitive or noncompetitive. Noncompetitive immunoassays are assays in which the amount of antigen is directly measured. In one preferred "sandwich" assay, for example, the anti- BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 antibodies can be bound directly to a solid substrate on which they are immobilized. These immobilized antibodies then capture BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 present in the test sample. BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 proteins thus immobilized are then bound by a labeling agent, such as a second BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 antibody bearing a label. Alternatively, the second antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second antibody is derived. The second or third antibody is typically modified with a detectable moiety, such as biotin, to which another molecule specifically binds, e.g., streptavidin, to provide a detectable moiety.

Competitive assay formats

In competitive assays, the amount of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein present in the sample is measured indirectly by measuring the amount of a known, added (exogenous) BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein displaced (competed away) from an anti- BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 antibody by the unknown BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein present in a

sample. In one competitive assay, a known amount of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein is added to a sample and the sample is then contacted with an antibody that specifically binds to BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein. The amount of exogenous BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein bound to the antibody is inversely proportional to the concentration of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein present in the sample. In a particularly preferred embodiment, the antibody is immobilized on a solid substrate. The amount of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein bound to the antibody may be determined either by measuring the amount of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 present in BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein/antibody complex, or alternatively by measuring the amount of remaining uncomplexed protein. The amount of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein may be detected by providing a labeled BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 molecule.

A hapten inhibition assay is another preferred competitive assay. In this assay the known BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein is immobilized on a solid substrate. A known amount of anti- BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 antibody is added to the sample, and the sample is then contacted with the immobilized BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD,

HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1. The amount of anti-BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 antibody bound to the known immobilized BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 is inversely proportional to the amount of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein present in the sample. Again, the amount of immobilized antibody may be detected by detecting either the immobilized fraction of antibody or the fraction of the antibody that remains in solution. Detection may be direct where the antibody is labeled or indirect by the subsequent addition of a labeled moiety that specifically binds to the antibody as described above.

Cross-reactivity determinations

Immunoassays in the competitive binding format can also be used for crossreactivity determinations. For example, a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein can be immobilized to a solid support. Proteins (e.g., BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 and homologs) are added to the assay that compete for binding of the antisera to the immobilized antigen. The ability of the added proteins to compete for binding of the antisera to the immobilized protein is compared to the ability of the BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein to compete with itself. The percent crossreactivity for the above proteins is calculated, using standard calculations. Those antisera with less than 10% crossreactivity with each of the added proteins listed above are selected and pooled. The cross-reacting antibodies are optionally removed from the pooled antisera by immunoabsorption with the added considered proteins, e.g., distantly related homologs.

The immunoabsorbed and pooled antisera are then used in a competitive binding immunoassay as described above to compare a second protein, thought to be perhaps an allele or polymorphic variant of a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1,

aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein, to the immunogen protein. In order to make this comparison, the two proteins are each assayed at a wide range of concentrations and the amount of each protein required to inhibit 50% of the binding of the antisera to the immobilized protein is determined. If the amount of the second protein required to inhibit 50% of binding is less than 10 times the amount of the BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein that is required to inhibit 50% of binding, then the second protein is said to specifically bind to the polyclonal antibodies generated to BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 immunogen.

Other assay formats

Western blot (immunoblot) analysis is used to detect and quantify the presence of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 in the sample. The technique generally comprises separating sample proteins by gel electrophoresis on the basis of molecular weight, transferring the separated proteins to a suitable solid support, (such as a nitrocellulose filter, a nylon filter, or derivatized nylon filter), and incubating the sample with the antibodies that specifically bind BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1. The anti-BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 antibodies specifically bind to the BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 on the solid support. These antibodies may be directly labeled or alternatively may be subsequently detected using labeled antibodies (e.g., labeled sheep anti-mouse antibodies) that specifically bind to the anti-BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 antibodies.

Other assay formats include liposome immunoassays (LIA), which use liposomes designed to bind specific molecules (e.g., antibodies) and release encapsulated

reagents or markers. The released chemicals are then detected according to standard techniques (*see* Monroe *et al., Amer. Clin. Prod. Rev.* 5:34-41 (1986)).

Reduction of non-specific binding

One of skill in the art will appreciate that it is often desirable to minimize non-specific binding in immunoassays. Particularly, where the assay involves an antigen or antibody immobilized on a solid substrate it is desirable to minimize the amount of non-specific binding to the substrate. Means of reducing such non-specific binding are well known to those of skill in the art. Typically, this technique involves coating the substrate with a proteinaceous composition. In particular, protein compositions such as bovine serum albumin (BSA), nonfat powdered milk, and gelatin are widely used with powdered milk being most preferred.

Labels

The particular label or detectable group used in the assay is not a critical aspect of the invention, as long as it does not significantly interfere with the specific binding of the antibody used in the assay. The detectable group can be any material having a detectable physical or chemical property. Such detectable labels have been well-developed in the field of immunoassays and, in general, most any label useful in such methods can be applied to the present invention. Thus, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include magnetic beads (e.g., DYNABEADS™), fluorescent dyes (e.g., fluorescein isothiocyanate, Texas red, rhodamine, and the like), radiolabels (e.g., ^{3}H, ^{125}I, ^{35}S, ^{14}C, or ^{32}P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic beads (e.g., polystyrene, polypropylene, latex, etc.).

The label may be coupled directly or indirectly to the desired component of the assay according to methods well known in the art. As indicated above, a wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions.

Non-radioactive labels are often attached by indirect means. Generally, a ligand molecule (e.g., biotin) is covalently bound to the molecule. The ligand then binds to

another molecules (e.g., streptavidin) molecule, which is either inherently detectable or covalently bound to a signal system, such as a detectable enzyme, a fluorescent compound, or a chemiluminescent compound. The ligands and their targets can be used in any suitable combination with antibodies that recognize BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein, or secondary antibodies that recognize anti- BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1.

The molecules can also be conjugated directly to signal generating compounds, e.g., by conjugation with an enzyme or fluorophore. Enzymes of interest as labels will primarily be hydrolases, particularly phosphatases, esterases and glycosidases, or oxidotases, particularly peroxidases. Fluorescent compounds include fluorescein and its derivatives, rhodamine and its derivatives, dansyl, umbelliferone, etc. Chemiluminescent compounds include luciferin, and 2,3-dihydrophthalazinediones, e.g., luminol. For a review of various labeling or signal producing systems that may be used, see U.S. Patent No. 4,391,904.

Means of detecting labels are well known to those of skill in the art. Thus, for example, where the label is a radioactive label, means for detection include a scintillation counter or photographic film as in autoradiography. Where the label is a fluorescent label, it may be detected by exciting the fluorochrome with the appropriate wavelength of light and detecting the resulting fluorescence. The fluorescence may be detected visually, by the use of electronic detectors such as charge coupled devices (CCDs) or photomultipliers and the like. Similarly, enzymatic labels may be detected by providing the appropriate substrates for the enzyme and detecting the resulting reaction product. Colorimetric or chemiluminescent labels may be detected simply by observing the color associated with the label. Thus, in various dipstick assays, conjugated gold often appears pink, while various conjugated beads appear the color of the bead.

Some assay formats do not require the use of labeled components. For instance, agglutination assays can be used to detect the presence of the target antibodies. In this case, antigen-coated particles are agglutinated by samples comprising the target antibodies. In this format, none of the components need be labeled and the presence of the target antibody is detected by simple visual inspection.

CELLULAR TRANSFECTION AND GENE THERAPY

The present invention provides the nucleic acids of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein for the transfection of cells *in vitro* and *in vivo*. These nucleic acids can be inserted into any of a number of well-known vectors for the transfection of target cells and organisms as described below. The nucleic acids are transfected into cells, *ex vivo* or *in vivo*, through the interaction of the vector and the target cell. The nucleic acid, under the control of a promoter, then expresses a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein of the present invention, thereby mitigating the effects of absent, partial inactivation, or abnormal expression of a BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 gene, particularly as it relates to cellular proliferation. The compositions are administered to a patient in an amount sufficient to elicit a therapeutic response in the patient. An amount adequate to accomplish this is defined as "therapeutically effective dose or amount."

Such gene therapy procedures have been used to correct acquired and inherited genetic defects, cancer, and other diseases in a number of contexts. The ability to express artificial genes in humans facilitates the prevention and/or cure of many important human diseases, including many diseases which are not amenable to treatment by other therapies (for a review of gene therapy procedures, *see* Anderson, *Science* 256:808-813 (1992); Nabel & Felgner, TIBTECH 11:211-217 (1993); Mitani & Caskey, TIBTECH 11:162-166 (1993); Mulligan, *Science* 926-932 (1993); Dillon, TIBTECH 11:167-175 (1993); Miller, *Nature* 357:455-460 (1992); Van Brunt, *Biotechnology* 6(10):1149-1154 (1998); Vigne, *Restorative Neurology and Neuroscience* 8:35-36 (1995); Kremer & Perricaudet, *British Medical Bulletin* 51(1):31-44 (1995); Haddada *et al.*, in *Current Topics in Microbiology and Immunology* (Doerfler & Böhm eds., 1995); and Yu *et al.*, *Gene Therapy* 1:13-26 (1994)).

PHARMACEUTICAL COMPOSITIONS AND ADMINISTRATION

Pharmaceutically acceptable carriers are determined in part by the particular composition being administered (e.g., nucleic acid, protein, modulatory compounds or transduced cell), as well as by the particular method used to administer the composition.

Accordingly, there are a wide variety of suitable formulations of pharmaceutical compositions of the present invention (*see, e.g., Remington's Pharmaceutical Sciences*, 17th ed., 1989). Administration can be in any convenient manner, e.g., by injection, oral administration, inhalation, transdermal application, or rectal administration.

Formulations suitable for oral administration can consist of (a) liquid solutions, such as an effective amount of the packaged nucleic acid suspended in diluents, such as water, saline or PEG 400; (b) capsules, sachets or tablets, each containing a predetermined amount of the active ingredient, as liquids, solids, granules or gelatin; (c) suspensions in an appropriate liquid; and (d) suitable emulsions. Tablet forms can include one or more of lactose, sucrose, mannitol, sorbitol, calcium phosphates, corn starch, potato starch, microcrystalline cellulose, gelatin, colloidal silicon dioxide, talc, magnesium stearate, stearic acid, and other excipients, colorants, fillers, binders, diluents, buffering agents, moistening agents, preservatives, flavoring agents, dyes, disintegrating agents, and pharmaceutically compatible carriers. Lozenge forms can comprise the active ingredient in a flavor, e.g., sucrose, as well as pastilles comprising the active ingredient in an inert base, such as gelatin and glycerin or sucrose and acacia emulsions, gels, and the like containing, in addition to the active ingredient, carriers known in the art.

The compound of choice, alone or in combination with other suitable components, can be made into aerosol formulations (i.e., they can be "nebulized") to be administered via inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like.

Formulations suitable for parenteral administration, such as, for example, by intraarticular (in the joints), intravenous, intramuscular, intradermal, intraperitoneal, and subcutaneous routes, include aqueous and non-aqueous, isotonic sterile injection solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic with the blood of the intended recipient, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. In the practice of this invention, compositions can be administered, for example, by intravenous infusion, orally, topically, intraperitoneally, intravesically or intrathecally. Parenteral administration and intravenous administration are the preferred methods of administration. The formulations of commends can be presented in unit-dose or multi-dose sealed containers, such as ampules and vials.

Injection solutions and suspensions can be prepared from sterile powders, granules, and tablets of the kind previously described. Cells transduced by nucleic acids for *ex vivo* therapy can also be administered intravenously or parenterally as described above.

The dose administered to a patient, in the context of the present invention should be sufficient to effect a beneficial therapeutic response in the patient over time. The dose will be determined by the efficacy of the particular vector employed and the condition of the patient, as well as the body weight or surface area of the patient to be treated. The size of the dose also will be determined by the existence, nature, and extent of any adverse side-effects that accompany the administration of a particular vector, or transduced cell type in a particular patient.

In determining the effective amount of the vector to be administered in the treatment or prophylaxis of conditions owing to diminished or aberrant expression of the BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, G6PD, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily member 1, or ERCC1 protein, the physician evaluates circulating plasma levels of the vector, vector toxicities, progression of the disease, and the production of anti-vector antibodies. In general, the dose equivalent of a naked nucleic acid from a vector is from about 1 μg to 100 μg for a typical 70 kilogram patient, and doses of vectors which include a retroviral particle are calculated to yield an equivalent amount of therapeutic nucleic acid.

For administration, compounds and transduced cells of the present invention can be administered at a rate determined by the LD-50 of the inhibitor, vector, or transduced cell type, and the side-effects of the inhibitor, vector or cell type at various concentrations, as applied to the mass and overall health of the patient. Administration can be accomplished via single or divided doses.

EXAMPLES

The following examples are offered to illustrate, but not to limit the claimed invention.

Example 1: Isolation Of Genes Which Cause Cell Cycle Arrest

A GFP C-terminal cDNA fusion library with a tetOff inducible gene expression system was constructed using standard techniques known to those of skill in the art. Clones from the library were used to transfect A549 cells. Transfected cells were then stained with cell tracker dyes to monitor the cell cycle. Cell tracker intensity correlated with

p21 expression. p21-induced arrested cells are also resistant to retrovirus infection. After transfection with the cDNA library, cells that stained more brightly with cells tracker dyes were identified as cell cycle arrested cells. Cycling cells were eliminated by transfection with a retrovirus encoding the diphtheria toxin alpha chain. Cycling cells are susceptible to retroviral infection, but cell cycle arrested cells are not. Cell tracker positive cells, i.e., cell cycle arrested cells, were sorted into 96 well plates and expanded with doxycycline (Dox) treatment. AlamarBlue, an oxidation-reduction indicator, was used to evaluate the proliferative effect of Dox on individual clones. AlamarBlue exhibits a spectrophotometrically measurable shift in color when reduced, e.g., within a proliferating cell. Clones that failed to proliferate in the presence of Dox were identified as clones encoding genes that had antiproliferative effects. Phenotype transfer into naïve A549 cells was performed with Dox-regulatable clones. The gene or gene fragment of interest was then amplified by RT-PCR.

Example 2: Identification of Antiproliferative Proteins

A549 cells were transfected with a clone containing a fragment of BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily 1, ERCC1, or a fragment thereof. The transfected cells were stained with a cell cycle tracker dye. The BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily 1, and ERCC1 transfected cells stained brightly with the cell cycle tracker dye, indicating that they were cell cycle arrested cells. Thus, BAP-1, NP95, FANCA, DDX9, IGF1R, UBE2V1, aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DDX21, ARK2, transmembrane 4 superfamily 1, and ERCC1 were identified as antiproliferative proteins.

Example 3: Assay for UBE2V1 Activity

UBE2V1 activity can be assessed using an *in vitro* ubiquitination assay as described in Sancho *et al., Mol. Cell. Biol.* 18(1):576 (1998). Briefly, UBE2V1 or a sample suspected of containing UBE2V1 is incubated with ^{125}I-ubiquitin at 37°C for 2 hours and conjugation of UBE2V1 to ^{125}I-ubiquitin is measured.

Example 8: Assay for Pyruvate Kinase Activity

Pyruvate kinase activity can be assessed according to the method described in Melo *et al.*, *Cell. Biochem. Func.* 16:99 (2001). Briefly, the rate of NADH oxidation at 30°C is measured in a coupled LDH assay system. The reaction mixture contains 50 mM Tris-HCl buffer at pH 7.5, 0.5 mM NADH, 10 mM KCl, 5 mM $MgSO_4$, 1 mM EDTA, 3 MM ADP, 0.5 mM DTT, 1U/ml LDH, and an appropriate amount of cellular extract. The reaction can be initiated by the addition of 2.5 mM phosphoenopyruvate. NADH oxidation can be follow using the molar extinction coefficient 6.22 x 10^3 M/cm at 340 nm. One unit of pyruvate kinase is the amount of enzyme sufficient to oxidize 1 μmol NADH per minute. Enzyme activity can be measured spectrophotometrically with a Gilford spectrophtometer coupled to a recorder.

Example 9: Assay for Glucose-6-phosphate Dehydrogenase Activity

G6PD activity can be measured according to the method described in Ho *et al.*, *Free Rad. Biol. Med.*, 29(2):156 (2000). Briefly, cell extracts are prepared and an appropriate amount of cell extract is suspended in 1 ml of assay buffer: 50 mM Tris-HCl, pH 8.0, 50 mM $MgCl_2$, 4 mM NADP+, and 4 mM glucose-6-phosphate. The reduction of NADP+ in the presence of glucose-6-phosphate is indicative of enzymatic activity. G6PD activity can be measured spetrophotometrically at 340 nm.

Example 11: Assay for DDX21 Activity

RNA helicase activity of DDX21 can be measured according to the method described in Valdez, *Eur. J. Biochem.* 267:6395 (2000). Briefly, two RNA substrates can be prepared by synthesizing RNA in the presence of [α-^{32}P]GTP and gel purifying the RNA. Denatured or boiled ssRNA is mixed with RNA helicase purified from cell extracts in an assay buffer containing 20 mM Hepes/KOH, pH 7.6, 2 mM DTT, 3 mM $MgCl_2$, 0.1 M KCl, 2 units RNase inhibitor, 100 fmoles ssRNA substrate, and 20-50 ng protein from cell extracts. The reaction is incubated at 30°C for 20 minutes. The reaction is terminated by the addition of a loading buffer containing 0.1 M Tris-Hcl, pH 7.4, 20 mM EDTA, 0.5% SDS, 0.1% NP40, 0.1% bromophenol blue, 0.1% xylene cyanol, 50% glycerol, and 0.2 mg/ml proteinase K. The terminated reaction is run out on a 10% SDS/polyacrylamide gel at 100 V at room temperature. Folded RNA is identified easily because it migrates more slowly on a gel than the ssRNA substrate.

Example 12: BAP-1 WT protein, protease mutants, siRNA and antisense functional hit are antiproliferative.

The BAP-1 functional hit identified in the retroviral screen is in the antisense orientation. (Figure 2). Expression of the functional hit in a tumor cell line, *e.g.*, A549 cells, or in untransformed cells, *e.g.*, HMEC or PrEc cells, was antiproliferative. (See, *e.g.*, Figures 3, and 34-35.)

Dominant negative mutants of BAP-1 were made by mutating residues in the protease domain. (See, *e.g.*, Figure 29.) Using two different assays, expression of BAP-1 wild-type and protease mutants was antiproliferative in tumor cell lines, *i.e.*, HeLa cells and H1299 cells. (See, *e.g.*, Figures 30-33). siRNA molecules derived from the BAP-1 nucleic aicd were shown to be antiproiferative in HeLa cells and H1299 cells. (See, *e.g.*, Figures 36-37.)

Example 13: BAP-1 is a ubiquitin protease.

GST-Bap-1 was expressed in and purified from SF9 cells. (See, *e.g.*, Figures 38-39.) Using a fluorogenic ubiuquiting cleavage assay, BAP-1 was shown to be an active ubiquitin protease, with a Km of 0.5 μM for the substrate UbAMC. (See, *e.g.*, Figures 40-42.) UbCHO was also demonstrated to be a specific inhibitor of BAP-1. (See, *e.g.*, Figure 43.)

Assays for ubiquitin hydrolase activity (*e.g.*, to assay BAP-1 activity) can also be performed as described in U.S. Patent No. 6,307,035 and Mayer and Wilkinson, *Biochemistry* 28:166(1989) using the glycine 76 ethyl ester of ubiquitin as a substrate. Peak areas can be integrated and normalized with respect to the ubiquitin standard.

Example 14: NP95 WT protein, ring finger mutants, siRNA and functional hit are antiproliferative.

The NP95 (G1-2635) functional hit (G1-2635) identified in the retroviral screen is in the sense orientation. (Figure 2). Expression of the functional hit in a tumor cell line, *e.g.*, A549 cells, or in untransformed cells, *e.g.*, HMEC or PrEc cells, was antiproliferative. (See, *e.g.*, Figures 6, and 44-45.) siRNA molecules derived from the NP-95 nucleic acid were shown to be antiproiferative in PrEc and HUVEC cells and H1299 cells. (See, *e.g.*, Figures 46-47, and 57.)

Using real time PCR analysis, NP95 mRNA expression was shown to be overexpressed in tumor tissue relative to normal tissue from the same patient. Increased

NP95 expression was demonstrated in breast, lung and prostate cancer. (See, *e.g.*, Figures 48-50.)

Dominant negative mutants of NP95 were made by mutating residues the RING finger domain. A RING finger deletion mutant, ΔRING was also construced. (See, *e.g.*, Figure 51.) Expression of NP95 wild-type and RING finger mutants was antiproliferative in a tumor cell lines, *i.e.*, HCT116 cells, and in primary cells, *i.e.*, HMEC and PrEc cells. (See, *e.g.*, Figures 52 and 55-56). Expression of NP95 wild-type and RING finger mutants was not antiproliferative in a second tumor cell lines, *i.e.*, A549 cells. (See, *e.g.*, Figure 53). However, expression of the NP95 ΔRING mutant rendered the A549 cells sensitive to treatment with Bleomycin. (See, *e.g.*, Figure 54).

Example 15: NP95 is a ubiquitin ligase.

Figure 58 depicts the biochemistry of ubiquitinylation. NP95 exhibits E3 ubiquitin ligase activity, and the RING domain of NP95 is required for that activity. (see, *e.g.*, Figures 59-60.) NP95 can be expressed and purified from SF9 cells for use in enzymatic assays. (See, *e.g.*, Figure 61.) A plate based assay for ubiquitin ligase activity is shown schematically in figure 62 and described in (*see, e.g.*, WO 01/75145). NP95 exhibits ubiquitin ligase activity in that assay system. (See, *e.g.*, Figure 63.)

It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.

WE CLAIM:

1. A method for identifying a compound that modulates cell cycle arrest, the method comprising the steps of:

(i) contacting a cell comprising a target polypeptide selected from the group consisting of BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DEAD/H box polypeptide 21 (DDX21), serine threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, or ERCC1, or fragment thereof with the compound, the target polypeptide encoded by a nucleic acid that hybridizes under stringent conditions to a nucleic acid encoding a polypeptide having an amino acid sequence a sequence selected from the group consisting of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, and 28; and

(ii) determining the chemical or phenotypic effect of the compound upon the cell comprising the target polypeptide or fragment thereof, thereby identifying a compound that modulates cell cycle arrest.

2. The method of claim 1, wherein the chemical or phenotypic effect is determined by measuring an activity selected from the group consisting of: helicase activity, receptor tyrosine kinase activity, ubiquitination, ligase, ubiquitin hydrolase activity, ubiquitin ligase activity, receptor binding activity, receptor cross-linking acitivity, protease, and endonuclease.

3. The method of claim1, wherein the chemical or phenotypic effect is determined by measuring cellular proliferation.

4. The method of claim 3, wherein the cell cycle arrest is measured by assaying DNA synthesis or fluorescent marker level.

5. The method of claim 4, wherein DNA synthesis is measured by ^{3}H thymidine incorporation, BrdU incorporation, or Hoescht staining.

6. The method of claim 4, wherein the fluorescent marker is selected from the group consisting of a cell tracker dye or green fluorescent protein.

7. The method of claim 1, wherein modulation is activation of cell cycle arrest.

8. The method of claim 1, wherein modulation is activation of cancer cell cycle arrest.

9. The method of claim 1, wherein the host cell is a cancer cell.

10. The method of claim 9, wherein the cancer cell is a breast, prostate, colon, or lung cancer cell.

11. The method of claim 9, wherein the cancer cell is a transformed cell line.

12. The method of claim 11, wherein the transformed cell line is PC3, H1299, MDA-MB-231, MCF7, A549, or HeLa.

13. The method of claim 9, wherein the cancer cell is p53 null or mutant.

14. The method of claim 9, wherein the cancer cell is p53 wild-type.

15. The method of claim 1, wherein the polypeptide is recombinant.

16. The method of claim 1, wherein the polypeptide is encoded by a nucleic acid comprising a sequence of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, or 27.

17. The method of claim 1, wherein the compound is an antibody.

18. The method of claim 1, wherein the compound is an antisense molecule.

19. The method of claim 1, wherein the compound is an RNAi molecule.

20. The method of claim 1, wherein the compound is a small organic molecule.

21. The method of claim 1, wherein the compound is a peptide.

22. The method of claim 21, wherein the peptide is circular.

23. A method for identifying a compound that modulates cell cycle arrest, the method comprising the steps of:

(i) contacting the compound with a target polypeptide selected from the group consisting of BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DEAD/H box polypeptide 21 (DDX21), serine threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, or ERCC1, or fragment thereof, the target polypeptide encoded by a nucleic acid that hybridizes under stringent conditions to a nucleic acid encoding a polypeptide having an amino acid sequence a sequence selected from the group consisting of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, and 28;

(ii) determining the physical effect of the compound upon the target polypeptide; and

(iii) determining the chemical or phenotypic effect of the compound upon a cell comprising the target polypeptide or fragment thereof, thereby identifying a compound that modulates cell cycle arrest.

24. A method of modulating cell cycle arrest in a subject, the method comprising the step of administering to the subject a therapeutically effective amount of a compound identified using the method of claim 1.

25. The method of claim 24, wherein the subject is a human.

26. The method of claim 25, wherein the subject has cancer.

27. The method of claim 24, wherein the compound is an antibody.

28. The method of claim 24, wherein the compound is an antisense molecule.

29. The method of claim 24, wherein the compound is an RNAi molecule.

30. The method of claim 24, wherein the compound is a small organic molecule.

31. The method of claim 24, wherein the compound is a peptide.

32. The method of claim 31, wherein the peptide is circular.

33. The method of claim 24, wherein the compound inhibits cancer cell proliferation.

34. A method of modulating cell cycle arrests in a subject, the method comprising the step of administering to the subject a therapeutically effective amount of a target polypeptide selected from the group consisting of BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DEAD/H box polypeptide 21 (DDX21), serine threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, or ERCC1, or fragment thereof, the target polypeptide encoded by a nucleic acid that hybridizes under stringent conditions to a nucleic acid encoding a polypeptide having an amino acid sequence a sequence selected from the group consisting of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, and 28.

35. A method of modulating cell cycle arrest in a subject, the method comprising the step of administering to the subject a therapeutically effective amount of a nucleic acid encoding a target polypeptide selected from the group consisting of BRCA-1-Associated Protein-1 (BAP-1), Nuclear Protein 95 (NP95), Fanconi anemia group A protein (FANCA), DEAD/H box polypeptide 9 (DDX9), insulin-like growth factor 1 receptor (IGF1R), ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), aldehyde dehydrogenase, pyruvate kinase, glucose-6-phosphate dehydrogenase, HCDR-3, DEAD/H box polypeptide 21 (DDX21), serine threonine kinase 15 (ARK2), transmembrane 4 superfamily member 1, or ERCC1, or fragment thereof, the nucleic

acid hybridizing under stringent conditions to a nucleic acid encoding a polypeptide having an amino acid sequence a sequence selected from the group consisting of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26 and 28.

SEQ ID NO:1
Size: 410
DNA--BAP-1

```
   1 gcccgttgtc tgtgtgtggg actgaggggc cccgggggcg gtgggggctc ccggtggggg
  61 cagcggtggg gagggagggc ctggacatgg cgctgagggg ccgcccccgcg ggaagatgaa
 121 taagggctgg ctggaagctgg agagcgaccc aggcctcttc accctgctcg tggaagattt
 181 cggtgtcaag gggtgcaag tggaggagat ctacgacctt cagagcaaat gtcagggccc
 241 tgtatatgga tttatcttcc tgttcaaatg gatcgaagag cgccggtccc ggcgaaaggt
 301 ctctaccttg gtggatgata cgtccgtgat tgatgatgat attgtgaata acatgttctt
 361 tgcccaccag ctgataccca actcttgtgc aactcatgcc ttgctgagcg tgctcctgaa
 421 ctgcagcagc gtggacctgg gacccaccct gagtcgcatg aaggacttca ccaagggttt
 481 cagccctgag agcaaaggat atgcgattgg caatgccccg gagttggcca aggcccataa
 541 tagccatgcc aggcccgagc cacgccacct ccctgagaag cagaatggcc ttagtgcagt
 601 gcggaccatg gaggcgttcc actttgtcag ctatgtgcct atcacaggcc ggctctttga
 661 gctggatggg ctgaaggtct accccattga ccatgggccc tgggggagg acgaggagtg
 721 gacagacaag gcccggcggg tcatcatgga gcgtatcggc ctcgccactg caggggagcc
 781 ctaccacgac atccgcttca acctgatggc agtggtgccc gaccgcagga tcaagtatga
 841 ggccaggctg catgtgctga aggtgaaccg tcagacagta ctagaggctc tgcagcagct
 901 gataagagta acacagccag agctgattca gacccacaag tctcaagagt cacagctgcc
 961 tgaggagtcc aagtcagcca gcaacaagtc cccgctggtg ctggaagcaa acagggcccc
1021 tgcagcctct gagggcaacc acacagatgg tgcagaggag gcggctggtt catgcgcaca
1081 agccccatcc cacagccctc ccaacaaacc caagctagtg gtgaagcctc caggcagcag
1141 cctcaatggg gttcacccca accccactcc cattgtccag cggctgccgg cctttctaga
1201 caatcacaat tatgccaagt cccccatgca ggaggaagaa gacctggcgg caggtgtggg
1261 ccgcagccga gttccagtcc gcccacccca gcagtactca gatgatgagg atgactatga
1321 ggatgacgag gaggatgacg tgcagaacac caactctgcc cttaggtata aggggaaggg
1381 aacagggaag ccaggggcat tgagcggttc tgctgatggg caactgtcag tgctgcagcc
1441 caacaccatc aacgtcttgg ctgagaagct caaagagtcc cagaaggacc tctcaattcc
1501 tctgtccatc aagactagca gcggggctgg gagtccggct gtggcagtgc ccacacactc
1561 gcagccctca cccaccccca gcaatgagag tacagacacg gcctctgaga tcggcagtgc
1621 tttcaactcg ccactgcgct cgcctatccg ctcagccaac ccgacgcggc cctccagccc
1681 tgtcacctcc cacatctcca aggtgctttt tggagaggat gacagcctgc tgcgtgttga
1741 ctgcatacgc tacaaccgtg ctgtccgtga tctgggtcct gtcatcagca caggcctgct
1801 gcacctggct gaggatgggg tgctgagtcc cctggcgctg acagagggtg ggaagggttc
1861 ctcgccctcc atcagaccaa tccaaggcag ccaggggtcc agcagcccag tggagaagga
1921 ggtcgtggaa gccacggaca gcagagagaa gacggggatg gtgaggcctg gcgagccctt
1981 gagtggggag aaatactcac ccaaggagct gctggcactg ctgaagtgtg tggaggctga
2041 gattgcaaac tatgaggcgt gcctcaagga ggaggtagag aagaggaaga agttcaagat
2101 tgatgaccag agaaggaccc acaactacga tgagttcatc tgcaccttta tctccatgct
2161 ggctcaggaa ggcatgctgg ccaacctagt ggagcagaac atctccgtgc ggcggcgcca
2221 aggggtcagc atcggccggc tccacaagca gcggaagcct gaccggcgga aacgctctcg
2281 cccctacaag gccaagcgcc agtgaggact gctggccctg actctgcagc ccactcttgc
2341 cgtgtggccc tcaccagggt ccttccctgc cccacttccc cttttcccag tattactgaa
2401 tagtcccagc tggagagtcc aggccctggg aatgggagga accaggccac attccttcca
2461 tcgtgccctg aggcctgaca cggcagatca gccccatagt gctcaggagg cagcatctgg
2521 agttggggca cagcgaggta ctgcagcttc ctccacagcc ggctgtggag cagcaggacc
2581 tggcccttct gcctgggcag cagaatatat attttaccta tcagagacat ctatttttct
2641 gggctccaac ccaacatgcc accatgttga cataagttcc tacctgacta tgctttctct
2701 cctaggagct gtcctggtgg gcccaggtcc ttgtatcatg ccacggtccc aactacaggg
2761 tcctagctgg gggcctgggt gggccctggg ctctgggccc tgctgctcta gccccagcca
2821 ccagcctgtc cctgttgtaa ggaagccagg tcttctctct tcattcctct taggagagtg
2881 ccaaactcag ggacccagca ctgggctggg ttgggagtag ggtgtcccag tggggttggg
2941 gtgagcaggc tgctgggatc ccatggcctg agcagagcat gtgggaactg ttcagtggcc
3001 tgtgaactgt cttccttgtt ctagccaggc tgttcaagac tgctctccat agcaaggttc
3061 tagggctctt cgccttcagt gttgtggccc tagctatggg cctaaattgg gctctaggtc
3121 tctgtccctg gcgcttgagg ctcagaagag cctctgtcca gcccctcagt attaccatgt
```

FIG. 1 (1/2)

3181 ctcccctcta ggggtagcag agacagggtt gcttatagga agctggcacc actcagctct
3241 tcctgctact ccagtttcct cagcctctgc aaggcactca gggtggggga cagcaggatc
3301 aagacaaccc gttggagccc ctgtgttcca gaggacctga tgccaagggg taatgggccc
3361 agcagtgcct ctggagccca ggccccaaca cagcccatg gcctctgcca gatggctttg
3421 aaaaaggtga tccaagcagg ccccttatc tgtacatagt gactgagtgg ggggtgctgg
3481 caagtgtggc agctgcctct gggctgagca cagcttgacc cctctagccc ctgtaaatac
3541 tggatcaatg aatgaataaa actctcctaa gaatctcctg agaaaaaaaa aaaaaaaaa

SEQ ID NO:2
Size: 729
PRT--BAP-1

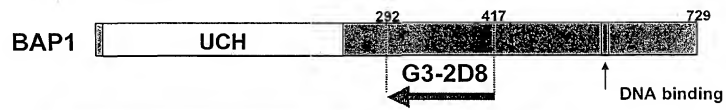
MNKGWLELESDPGLFTLLVEDFGVKGVQVEEIYDLQSKCQGPVYGFIFLFKWIEERRSRRKVSTLVDDTSVIDDD
IVNNMFFAHQLIPNSCATHALLSVLLNCSSVDLGPTLSRMKDFTKGFSPESKGYAIGNAPELAKAHNSHARPEPR
HLPEKQNGLSAVRTMEAFHFVSYVPITGRLFELDGLKVYPIDHGPWGEDEEWTDKARRVIMERIGLATAGEPYHD
IRFNLMAVVPDRRIKYEARLHVLKVNRQTVLEALQQLIRVTQPELIQTHKSQESQLPEESKSASNKSPLVLEANR
APAASEGNHTDGAEEAAGSCAQAPSHSPPNKPKLVVKPPGSSLNGVHPNPTPIVQRLPAFLDNHNYAKSPMQEEE
DLAAGVGRSRVPVRPPQQYSDDEDDYEDDEEDDVQNTNSALRYKGKGTGKPGALSGSADGQLSVLQPNTINVLAE
KLKESQKDLSIPLSIKTSSGAGSPAVAVPTHSQPSPTPSNESTDTASEIGSAFNSPLRSPIRSANPTRPSSPVTS
HISKVLFGEDDSLLRVDCIRYNRAVRDLGPVISTGLLHLAEDGVLSPLALTEGGKGSSPSIRPIQGSQGSSSPVE
KEVVEATDSREKTGMVRPGEPLSGEKYSPKELLALLKCVEAEIANYEACLKEEVEKRKKFKIDDQRRTHNYDEFI
CTFISMLAQEGMLANLVEQNISVRRRQGVSIGRLHKQRKPDRRKRSRPYKAKRQ

FIG. 1 (2/2)

WO 03/088910

G3-2D8 / BRCA1-Associated Protein-1 (BAP1)

The G3-2D8 sequence is identical to BRCA1-Associated Protein-1 (BAP1), 729aa
Orientation: Antisense



C-teminus of GFP

G3-2D8

GASTTCGTCACCCCCGCCGGGCATCACTCTCGGCATGGACGAGCTGTACAAGGAGGAGGCCGCCAAGGCN
E F V T A A G I T L G M D E L Y K E E A A K A

378 bp insert

GGTGGCAGCGGTGGCT *CCAGTGTGCTGGAAAG* CTAAGGGCAGAGTTGGTGTTCTGCACGTCATCCTC
G G S G G S S Y L E S In frame stop

TCGTCATCCTCATAGTCATCCTCATCATCTGAGTACTGCTGGGGTGGGCGGACTGGAACTCGGCTGCGG
CCCACACCTGCCGCCAGGTCTTCTTCCTCCTGCATGGGGGACTTGGCATAATTGTGATTGTCTAGAAAG
GCCGGCAGCCGCTGGACAATGGGAGTGGGGTTGGGGTGAACCCCATTGAGGCTGCTGCCTGGAGGCT
ACCACTAGCTTGGGTTTGTTGGGAGGGCTGTGGGATGGGGCTTGTGCGCATGAACCAGCCGCCTCCTC
GCACCATCTGTGTGGTTGCCCTCAGAGGCTGCAGGGGCCCTGTTTGCTTCCAGCACCAGCGGGGAC *CTT*
TCCAGCACAGTGG

CCAGTGTGCTGGAAAG,
CTTTCCAGCACAGTGG = BstXI linker

UCH(4-216): Ubiquitin carboxyl-terminal hydrolase, family 1,
DNA binding (625-640): 7kD DNA-binding domain

FIG. 2

PCT/US03/11867

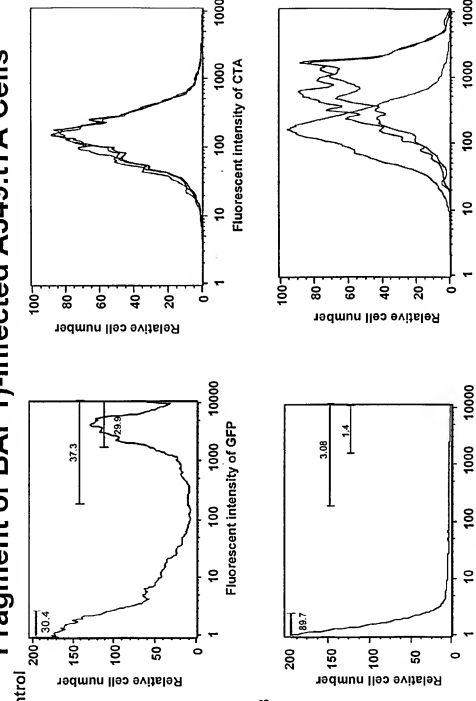

Cell Tracker Analysis of G3-2D8 (The Antisense Fragment of BAP1)-Infected A549.tTA Cells
GFP control
G3-2D8
Relative cell number
Fluorescent intensity of GFP
Fluorescent intensity of CTA
30.4
37.3
29.9
89.7
3.08
1.4
GFP negative cells
GFP positive cells
GFP highly positive cells

FIG. 3

SEQ ID NO:3

Size: 437
DNA--NP95

```
   1 CGACTCCTTA GAGCATGGCA TGGCTCAGAG GTGCTGGTAA AACTGATGGG GGTTTTTGCT
  61 GTCCCTCCCC TCAGCGCCGA CACCATGTGG ATCCAGGTTC GGACCATGGA CGGGAGGCAG
 121 ACCCACACGG TGGACTCGCT GTCCAGGCTG ACCAAGGTGG AGGAGCTGAG GCGGAAGATC
 181 CAGGAGCTGT TCCACGTGGA GCCAGGCCTG CAGAGGCTGT TCTACAGGGG CAAACAGATG
 241 GAGGACGGCC ATACCCTCTT CGACTACGAG GTCCGCCTGA ATGACACCAT CCAGCTCCTG
 301 GTCCGCCAGA GCCTCGTGCT CCCCCACAGC ACCAAGGAGC GGGACTCCGA GCTCTCCGAC
 361 ACCGACTCCG GCTGCTGCCT GGGCCAGAGT GAGTCAGACA AGTCCTCCAC CCACGGCGAG
 421 GCGGCCGCCG AGACTGACAG CAGGCCAGCC GATGAGGACA TGTGGGATGA GACGGAATTG
 481 GGGCTGTACA AGGTCAATGA GTACGTCGAT GCTCGGGACA CGAACATGGG GGCGTGGTTT
 541 GAGGCGCAGG TGGTCAGGGT GACGCGGAAG GCCCCCTCCC GGGACGAGCC CTGCAGCTCC
 601 ACGTCCAGGC CGGCGCTGGA GGAGGACGTC ATTTACCACG TGAAATACGA CGACTACCCG
 661 GAGAACGGCG TGGTCCAGAT GAACTCCAGG GACGTCCGAG CGCGCGCCCG CACCATCATC
 721 AAGTGGCAGG ACCTGGAGGT GGGCCAGGTG GTCATGCTCA ACTACAACCC CGACAACCCC
 781 AAGGAGCGGG GCTTCTGGTA CGACGCGGAG ATCTCCAGGA AGCGCGAGAC CAGGACGGCG
 841 CGGGAACTCT ACGCCAACGT GGTGCTGGGG GATGATTCTC TGAACGACTG TCGGATCATC
 901 TTCGTGGACG AAGTCTTCAA GATTGAGCGG CCGGGTGAAG GGAGCCCCAT GGTTGACAAC
 961 CCCATGAGAC GGAAGAGCGG GCCGTCCTGC AAGCACTGCA AGGACGACGT GAACAGACTC
1021 TGCCGGGTCT GCGCCTGCCA CCTGTGCGGG GGCCGGCAGG ACCCCGACAA GCAGCTCATG
1081 TGCGATGAGT GCGACATGGC CTTCCACATC TACTGCCTGG ACCCGCCCCT CAGCAGTGTT
1141 CCCAGCGAGG ACGAGTGGTA CTGCCCTGAG TGCCGGAATG ATGCCAGCGA GGTGGTACTG
1201 GCGGGAGAGC GGCTGAGAGA GAGCAAGAAG AAGGCGAAGA TGGCCTCGGC CACATCGTCC
1261 TCACAGCGGG ACTGGGGCAA GGGCATGGCC TGTGTGGGCC GCACCAAGGA ATGTACCATC
1321 GTCCCGTCCA ACCACTACGG ACCCATCCCG GGGATCCCCG TGGGCACCAT GTGGCGGTTC
1381 CGAGTCCAGG TCAGCGAGTC GGGTGTCCAT CGGCCCCACG TGGCTGGCAT ACACGGCCGG
1441 AGCAACGACG GAGCGTACTC CCTAGTCCTG GCGGGGGGCT ATGAGGATGA CGTGGACCAT
1501 GGGAATTTTT TCACATACAC GGGTAGTGGT GGTCGAGATC TTTCCGGCAA CAAGAGGACC
1561 GCGGAACAGT CTTGTGATCA GAAACTCACC AACACCAACA GGGCGCTGGC TCTCAACTGC
1621 TTTGCTCCCA TCAATGACCA AGAAGGGGCC GAGGCCAAGG ACTGGCGGTC GGGGAAGCCG
1681 GTCAGGGTGG TGCGCAATGT CAAGGGTGGC AAGAATAGCA AGTACGCCCC CGCTGAGGGC
1741 AACCGCTACG ATGGCATCTA CAAGGTTGTG AAATACTGGC CCGAGAAGGG GAAGTCCGGG
1801 TTTCTCGTGT GGCGCTACCT TCTGCGGAGG GACGATGATG AGCCTGGCCC TTGGACGAAG
1861 GAGGGGAAGG ACCGGATCAA GAAGCTGGGG CTGACCATGC AGTATCCAGA AGGCTACCTG
1921 GAAGCCCTGG CCAACCGAGA GCGAGAGAAG GAGAACAGCA AGAGGGAGGA GGAGGAGCAG
1981 CAGGAGGGGG GCTTCGCGTC CCCCAGGACG GGCAAGGGCA AGTGGAAGCG GAAGTCGGCA
2041 GGAGGTGGCC CGAGCAGGGC CGGGTCCCCG CGCCGGACAT CCAAGAAAAC CAAGGTGGAG
2101 CCCTACAGTC TCACGGCCCA GCAGAGCAGC CTCATCAGAG AGGACAAGAG CAACGCCAAG
2161 CTGTGGAATG AGGTCCTGGC GTCACTCAAG GACGGCCGG CGAGCGGCAG CCCGTTCCAG
2221 TTGTTCCTGA GTAAAGTGGA GGAGACGTTC CAGTGTATCT GCTGTCAGGA GCTGGTGTTC
2281 CGGCCCATCA CGACCGTGTG CCAGCACAAC GTGTGCAAGG ACTGCCTGGA CAGATCCTTT
2341 CGGGCACAGG TGTTCAGCTG CCCTGCCTGC CGCTACGACC TGGGCCGCAG CTATGCCATG
2401 CAGGTGAACC AGCCTCTGCA GACCGTCCTC AACCAGCTCT TCCCCGGCTA CGGCAATGGC
2461 CGGTGATCTC CAAGCACTTC TCGACAGGCG TTTTGCTGAA AACGTGTCGG AGGGCTCGTT
2521 CATCGGCACT GATTTGTTTC TTAGTGGGCT TAACTTAAAC AGGTAGTGTT TCCTCCGTTC
2581 CCTAAAAAGG TTTGTCTTCC TTTTTTTTTA TTTTTATTTT TCAAATCTAT ACATTTTCAG
2641 GAATTTATGT ATTCTGGCTA AAAGTTGGAC TTCTCAGTAT TGTGTTTAGT TCTTTGAAAA
2701 CATAAAAGCC TGCAATTTCT CGACAAAACA ACACAAGATT TTTTAAAGAT GGAATCAGAA
2761 ACTACGTGGT GTGGAGGCTG TTGATGTTTC TGGTGTCAAG TTCTCAGAAG TTGCTGCCAC
2821 CAACTCTTTA AGAAGGCGAC AGGATCAGTC CTTCTCTAGG GTTCTGGCCC CCAAGGTCAG
2881 AGCAAGCATC TTCCTGACAG CATTTTGTCA TCTAAAGTCC AGTGACATGG TTCCCCGTGG
2941 TGGCCCGTGG CAGCCCGTGG CATGGCGTGG CTCAGCTGTC TGTTGAAGTT GTTGCAAGGA
3001 AAAGAGGAAA CATCTCGGGC CTAGTTCAAA CCTTTGCCTC AAAGCCATCC CCCACCAGAC
3061 TGCTTAGCGT CTGAGATCCG CGTGAAAAGT CCTCTGCCCA CGAGAGCAGG GAGTTGGGGC
3121 CACGCAGAAA TGGCCTCAAG GGGACTCTGC TCCACGTGGG GCCAGGCGTG TGACTGACGC
```

FIG. 4 (1/2)

```
3181 TGTCCGACGA AGGCGGCCAC GGACGGACGC CAGCACACGA AGTCACGTGC AAGTGCCTTT
3241 GATTCGTTCC TTCTTTCTAA AGACGACAGT CTTTGTTGTT AGCACTGAAT TATTGAAAAT
3301 GTCAACCAGA TTCTAGAAAC TGCGGTCATC CAGTTCTTCC TGACACCGGA TGGGTGCTTG
3361 GGAACCGTTT GAGCCTTATA GATCATTTAC ATTCAATTTT TTTAACTCAG CAAGTGAGAA
3421 CTTACAAGAG GGTTTTTTTT TAATTTTTTT TTCTCTTAAT GAACACATTT TCTAAATGAA
3481 TTTTTTTTGT AGTTACTGTA TATGTACCAA GAAAGATATA ACGTTAGGGT TTGGTTGTTT
3541 TTGTTTTTGT ATTTTTTTTC TTTTGAAAGG GTTTGTTAAT TTTTCTAATT TTACCAAAGT
3601 TTGCAGCCTA TACCTCAATA AAACAGGGAT ATTTTAAATC ACATACCTGC AGACAAACTG
3661 GAGCAATGTT ATTTTTAAAG GGTTTTTTTC ACCTCCTTAT TCTTAGATTA TTAATGTATT
3721 AGGGAAGAAT GAGACAATTT TGTGTAGGCT TTTTCTAAAG TCCAGTACTT TGTCCAGATT
3781 TTAGATTCTC AGAATAAATG TTTTTCACAG ATTGAAAAAA AAAAAAA
```

SEQ ID NO:4
Size: 135
PRT--NP95

```
MWIQVRTMDGRQTHTVDSLSRLTKVEELRRKIQELFHVEPGLQRLFVRGKQMEDGHTLFDYEVRLNDTIQLLVRQ
SLVLPHSTKERDSELSDTDSGCCLGQSESDKSSTHGEAAAETDSRPADEDMWDETELGLYKVNEYVDARDTNMGA
WFEAQVVRVTRKAPSRDEEPCSSTSRPALEEDVIYHVKYDDYPENGVVQMNSRDVRARARTIIKWQDLEVGQVVML
NYNPDNPKERGFWYDAEISRKRETRTARELYANVVLGDDSLNDCRIIFVDEVFKIERPGEGSPMVDNPMRRKSGP
SCKHCKDDVNRLCRVCACHLCGGRQDPDKQLMCDECDMAPHIYCLDPPLSSVPSEDEWYCPECRNDASEVVLAGE
RLRESKKKAKMASATSSSQRDWGKGMACVGRTKECTIVPSNHVGPIPGIPVGTMWRFRVQVSESGVHRPHVAGIH
GRSNDGAYSLVLAGGYEDDVDHGNFFTYTGSGGRDLSGNKRTAEQSCDQKLTNTNRALALNCFAPINDQEGAEAK
DWRSGKPVRVVRNVKGGKNSKYAPAEGNRVDGIYKVVKYWPEKGKSGFLVWRYLLRDDDEPGPWTKEGKDRIKK
LGLTMQYPEGYLEALANRERSKENSKREEEEQQEGGFASPRTGKGKWKRKSAGGGPSRAGSPRRTSKKTKVEPYS
LTAQQSSLIREDKSNAKLWNEVLASLKDRPASGSPFQLPLSKVEETFQCICCQELVFRPITTVCQHNVCKDCLDR
SFRAQVFSCPACRYDLGRSYAMQVNQPLQTVLNQLFPGYGNGR
```

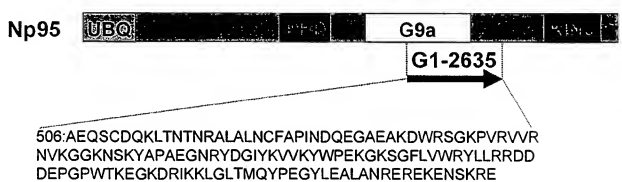
FIG. 4 (2/2)

G1-2635 / Np95

The G1-2635 sequence is identical to a nuclear zinc finger protein, Np95, 793aa

Orientation of cDNA: Sense

Pfam HMM search was done at the Washington University web site



UBQ(14-89): Ubiquitin like domain,
PHD(330-379): PHD-Zn finger, It could be important for the assembly or activity of multicomponent complexes
G9a(427-599): It is found in a nuclear protein associated with cell proliferation
RING(737-775): Zinc finger, C3HC4 type (RING finger), E3 ubiquitin-protein ligase activity is intrinsic to the RING domain of c-Cbl and is likely to be a general function of this domain; Various RING fingers exhibit binding to E2 ubiquitin-conjugating enzymes

FIG. 5

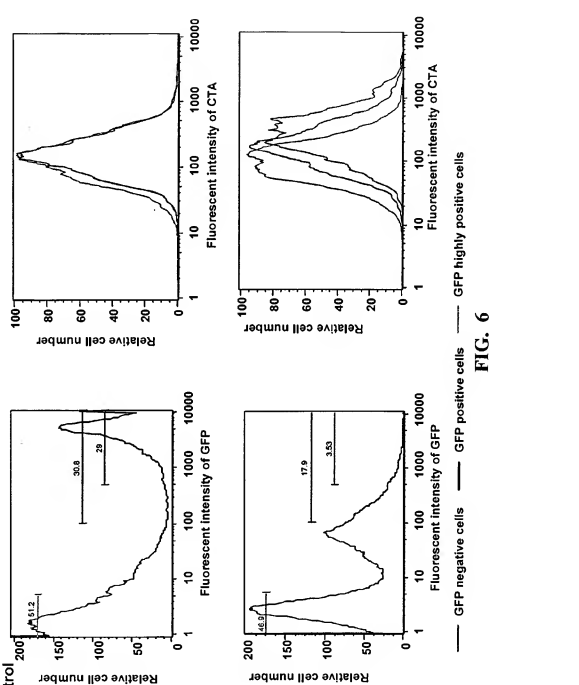
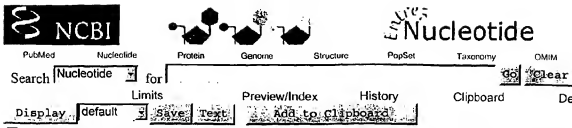

Cell Tracker Analysis of G1-2635 (The Fragment of Np95)-Infected A549.tTA Cells
GFP control
G1-2635
Relative cell number
Fluorescent intensity of GFP
Fluorescent intensity of CTA
51.2
30.8
29
46.9
17.9
3.53
GFP negative cells
GFP positive cells
GFP highly positive cells

FIG. 6

FIG. 7 (1/5)



LOCUS NM_000135 5503 bp mRNA linear PRI 05-JUL-2001
DEFINITION Homo sapiens Fanconi anemia, complementation group A (FANCA), mRNA.
ACCESSION NM_000135
VERSION NM_000135.1 GI:4503654
KEYWORDS .

FANC A

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5503)
AUTHORS Pronk JC, Gibson RA, Savoia A, Wijker M, Morgan NV, Melchionda S, Ford D, Temtamy S, Ortega JJ, Jansen S and et al.
TITLE Localisation of the Fanconi anaemia complementation group A gene to chromosome 16q24.3
JOURNAL Nat. Genet. 11 (3), 338-340 (1995)
MEDLINE 96042586
PUBMED 7581462
REFERENCE 2 (bases 1 to 5503)
AUTHORS Lo Ten Foe,J.R., Rooimans,M.A., Bosnoyan-Collins,L., Alon,N., Wijker,M., Parker,L., Lightfoot,J., Carreau,M., Callen,D.F., Savoia,A., Cheng,N.C., Van Berkel,C.G.M., Strunk,M.H.P., Gille,J.J.P., Pals,G., Kruyt,F.A.E., Pronk,J.C., Arwert,F., Buchwald,M. and Joenje,H.
TITLE Expression cloning of a cDNA for the major Fanconi anaemia gene, FAA
JOURNAL Nat. Genet. 14 (3), 320-323 (1996)
MEDLINE 97051928
REFERENCE 3 (bases 1 to 5503)
AUTHORS Ianzano L, D'Apolito M, Centra M, Savino M, Levran O, Auerbach AD, Cleton-Jansen AM, Doggett NA, Pronk JC, Tipping AJ, Gibson RA, Mathew CG, Whitmore SA, Apostolou S, Callen DF, Zelante L and Savoia A.
TITLE The genomic organization of the Fanconi anemia group A (FAA) gene
JOURNAL Genomics 41 (3), 309-314 (1997)
MEDLINE 97312685
PUBMED 9169126
REFERENCE 4 (bases 1 to 5503)
AUTHORS Joenje H, Oostra AB, Wijker M, di Summa FM, van Berkel CG, Rooimans MA, Ebell W, van Weel M, Pronk JC, Buchwald M and Arwert F.
TITLE Evidence for at least eight Fanconi anemia genes
JOURNAL Am. J. Hum. Genet. 61 (4), 940-944 (1997)
MEDLINE 98018453
PUBMED 9382107
REFERENCE 5 (bases 1 to 5503)
AUTHORS Kupfer GM, Naf D, Suliman A, Pulsipher M and D'Andrea AD.
TITLE The Fanconi anaemia proteins, FAA and FAC, interact to form a

FIG. 7 (2/5)

```
            nuclear complex
  JOURNAL   Nat. Genet. 17 (4), 487-490 (1997)
  MEDLINE   98061104
   PUBMED   9398857
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from X99226.1.
FEATURES             Location/Qualifiers
     source          1..5503
                     /organism="Homo sapiens"
                     /isolate="healthy control"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3"
                     /clone="D"
                     /cell_line="HSC93"
                     /cell_type="lymphoblastoid"
                     /clone_lib="pREP4"
     gene            1..5503
                     /gene="FANCA"
                     /note="FA; FA1; FAA; FACA; FANCH"
                     /db_xref="LocusID:2175"
                     /db_xref="MIM:227650"
                     /db_xref="MIM:603468"
     CDS             32..4399
                     /gene="FANCA"
                     /function="acts with other genes to control FA pathway"
                     /note="Fanconi anemia, complementation group H"
                     /codon_start=1
                     /db_xref="LocusID:2175"
                     /db_xref="MIM:227650"
                     /db_xref="MIM:603468"
                     /product="Fanconi anemia, complementation group A"
                     /protein_id="NP_000126.1"
                     /db_xref="GI:4503655"
                     /translation="MSDSWVPNSASGQDPGGRRRAWAELLAGRVKREKYNPERAQKLK
                     ESAVRLLRSHQDLNALLLEVEGPLCKKLSLSKVIDCDSSEAYANHSSSFIGSALQDQA
                     SRLGVPVGILSAGMVASSVGQICTAPAETSHPVLLTVEQRKKLSSLLEFAQYLLAHSM
                     FSRLSFCQELWKIQSSLLLEAVWHLHVQGIVSLQELLESHPDMHAVGSWLFRNLCCLC
                     EQMEASCQHADVARAMLSDFVQMFVLRGFQKNSDLRRTVEPEKMPQVTVDVLQRMLIF
                     ALDALAAGVQEESSTHKIVRCWFGVFSGHTLGSVISTDPLKRFFSHTLTQILTHSPVL
                     KASDAVQMQREWSFARTHPLLTSLYRRLFVMLSAEELVGHLQEVLETQEVHWQRVLSF
                     VSALVVCFPEAQQLLEDWVARLMAQAFESCQLDSMVTAFLVVRQAALEGPSAFLSYAD
                     WFKASFGSTRGYHGCSKKALVFLFTFLSELVPFESPRYLQVHILHPPLVPSKYRSLLT
                     DYISLAKTRLADLKVSIENMGLYEDLSSAGDITEPHSQALQDVEKAIMVFEHTGNIPV
                     TVMEASIFRRPYYVSHFLPALLTPRVLPKVPDSRVAFIESLKRADKIPPSLYSTYCQA
                     CSAAEEKPEDAALGVRAEPNSAEEPLGQLTAALGELRASMTDPSQRDVISAQVAVISE
                     RLRAVLGHNEDDSSVEISKIQLSINTPRLEPREHIAVDLLLTSFCQNLMAASSVAPPE
                     RQGPWAALFVRTMCGRVLPAVLTRLCQLLRHQGPSLSAPHVLGLAALAVHLGESRSAL
                     PEVDVGPPAPGAGLPVPALFDSLLTCRTRDSLFFCLKFCTAAISYSLCKFSSQSRDTL
                     CSCLSPGLIKKFQFLMFRLFSEARQPLSEEDVASLSWRPLHLPSADWQRAALSLWTHR
                     TFREVLKEEDVHLTYQDWLHLELEIQPEADALSDTERQDFHQWAIHEHFLPESSASGG
                     CDGDLQAACTILVNALMDFHQSSRSYDHSENSDLVFGGRTGNEDIISRLQEMVADLEL
                     QQDLIVPLGHTPSQEHFLFEIFRRRLQALTSGWSVAASLQRQRELLMYKRILLRLPSS
                     VLCGSSFQAEQPITARCEQFFHLVNSEMRNFCSHGGALTQDITAHFFRGLLNACLRSR
                     DPSLMVDFILAKCQTKCPLILTSALVWWPSLEPVLLCRWRRHCQSPLPRELQKLQEGR
                     QFASDFLSPEAASPAPNPDWLSAAALHFAIQQVREENIRKQLKKLDCEREELLVFLFF
                     FSLMGLLSSHLTSNSTTDLPKAFHVCAAILECLEKRKISWLALFQLTESDLRLGRLLL
                     RVAPDQHTRLLPFAFYSLLSYFHEDAAIREEAFLHVAVDMYLKLVQLFVAGDTSTVSP
                     PAGRSLELKGQGNPVELITKARLFLLQLIPRCPKKSFSHVAELLADRGDCDPEVSAAL
                     QSRQQAAPDADLSQEPHLF"
```

FIG. 7 (3/5)

```
     variation       48
                     /allele="A"
                     /allele="T"
                     /db_xref="dbSNP:1800282"
     variation       1174
                     /allele="G"
                     /allele="T"
                     /db_xref="dbSNP:1800331"
     variation       1321
                     /allele="A"
                     /allele="G"
                     /db_xref="dbSNP:1800332"
     variation       complement(1532)
                     /allele="C"
                     /allele="T"
                     /db_xref="dbSNP:2239359"
     variation       3214
                     /allele="C"
                     /allele="T"
                     /db_xref="dbSNP:1800346"
     variation       3685
                     /allele="A"
                     /allele="G"
                     /db_xref="dbSNP:1800358"
     variation       4553
                     /allele="A"
                     /allele="G"
                     /db_xref="dbSNP:1230"
BASE COUNT     1208 a   1527 c   1492 g   1276 t
ORIGIN
        1 agccgccgcc ggggctgtag gcgccaaggc catgtccgac tcgtgggtcc cgaactccgc
       61 ctcgggccag gacccagggg gccgccggag ggcctgggcc gagctgctgg cgggaagggt
      121 caagagggaa aaatataatc ctgaaagggc acagaaatta aaggaatcag ctgtgcgcct
      181 cctgcgaagc catcaggacc tgaatgccct tttgcttgag gtagaaggtc cactgtgtaa
      241 aaaattgtct ctcagcaaag tgattgactg tgacagttct gaggcctatg ctaatcattc
      301 tagttcattt ataggctctg ctttgcagga tcaagcctca aggctggggg ttcccgtggg
      361 tattctctca gccgggatgg ttgcctctag cgtgggacag atctgcacgg ctccagcgga
      421 gaccagtcac cctgtgctgc tgactgtgga gcagagaaag aagctgtctt ccctgttaga
      481 gtttgctcag tatttattgg cacacagtat gttctcccgt ctttccttct gtcaagaatt
      541 atggaaaata cagagttctt tgttgcttga agcggtgtgg catcttcacg tacaaggcat
      601 tgtgagcctg caagagctgc tggaaagcca tcccgacatg catgctgtgg gatcgtggct
      661 cttcaggaat ctgtgctgcc tttgtgaaca gatggaagca tcctgccagc atgctgacgt
      721 cgccagggcc atgctttctg attttgttca aatgtttgtt ttgaggggat ttcagaaaaa
      781 ctcagatctg agaagaactg tggagcctga aaaaatgccg caggtcacgg ttgatgtact
      841 gcagagaatg ctgatttttg cacttgacgc tttggctgct ggagtacagg aggagtcctc
      901 cactcacaag atcgtgaggt gctggttcgg agtgttcagt ggacacacgc ttggcagtgt
      961 aatttccaca gatcctctga agaggttctt cagtcatacc ctgactcaga tactcactca
     1021 cagccctgtg ctgaaagcat ctgatgctgt tcagatgcag agagagtgga gctttgcgcg
     1081 gacacaccct ctgctcacct cactgtaccg caggctcttt gtgatgctga gtgcagagga
     1141 gttggttggc catttgcaag aagttctgga aacgcaggag gttcactggc agagagtgct
     1201 ctcctttgtg tctgccctgg ttgtctgctt tccagaagcg cagcagctgc ttgaagactg
     1261 ggtggcgcgt ttgatggccc aggcattcga gagctgccag ctggacagca tggtcactgc
     1321 gttcctggtt gtgcgccagg cagcactgga gggcccctct gcgttcctgt catatgcaga
     1381 ctggttcaag gcctcctttg ggagcacacg aggctaccat ggctgcagca agaaggccct
     1441 ggtcttcctg tttacgttct tgtcagaact cgtgcctttt gagtctcccc ggtacctgca
     1501 ggtgcacatt ctccacccac ccctggttcc cagcaagtac cgctccctcc tcacagacta
     1561 catctcattg gccaagacac ggctggccga cctcaaggtt tctatagaaa acatgggact
     1621 ctacgaggat ttgtcatcag ctggggacat tactgagccc cacagccaag ctcttcagga
     1681 tgttgaaaag gccatcatgg tgtttgagca tacggggaac atcccagtca ccgtcatgga
     1741 ggccagcata ttcaggaggc cttactacgt gtcccacttc ctccccgccc tgctcacacc
```

FIG. 7 (4/5)

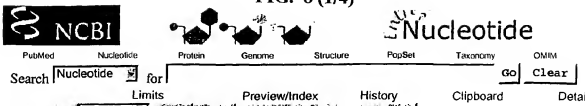
```
1801 tcgagtgctc cccaaagtcc ctgactcccg tgtggcgttt atagagtctc tgaagagagc
1861 agataaaatc ccccatctc tgtactccac ctactgccag gcctgctctg ctgctgaaga
1921 gaagccagaa gatgcagccc tgggagtgag ggcagaaccc aactctgctg aggagccctt
1981 gggacagctc acagctgcac tgggagagct gagagcctcc atgacagacc ccagccagcg
2041 tgatgttata tcggcacagg tggcagtgat ttctgaaaga ctgagggctg tcctgggcca
2101 caatgaggat gacagcagcg ttgagatatc aaagattcag ctcagcatca acacgccgag
2161 actggagcca cgggaacaca ttgctgtgga cctcctgctg acgtctttct gtcagaacct
2221 gatggctgcc tccagtgtcg ctccccgga gaggcagggt ccctgggctg ccctcttgct
2281 gaggaccatg tgtggacgtg tgctccctgc agtgctcacc cggctctgcc agctgctccg
2341 tcaccagggc ccgagcctga gtgccccaca tgtgctgggg ttggctgccc tggcccgtca
2401 cctgggtgag tccaggtctg cgctcccaga ggtggatgtg ggtcctcctg cacctggtgc
2461 tggccttcct gtcctcgcgc tctttgacag cctcctgacc tgtaggacga gggattcctt
2521 gttcttctgc ctgaaatttt gtacagcagc aatttcttac tctctctgca agtttttcttc
2581 ccagtcacga gatactttgt gcagctgctt atctccaggc cttattaaaa agtttcagtt
2641 cctcatgttc agattgttct cagaggcccg acagcctctt tctgaggagg acgtagccag
2701 cctttcctgg agacccttgc accttccttc tgcagactgg cagagagctg ccctctctct
2761 ctggacacac agaaccttcc gagaggtgtt gaaagaggaa gatgttcact taacttacca
2821 agactggtta cacctggagc tggaaattca acctgaagct gatgctcttt cagatactga
2881 acggcaggac ttccaccagt ggggcgatcca tgagcacttt ctccctgagt cctcggcttc
2941 aggggcgctgt gacggagacc tgcaggctgc gtgtaccatt cttgtcaacg cactgatgga
3001 tttccaccaa agctcaagga gttatgacca ctcagaaaat tctgatttgg tctttggtgg
3061 ccgcacagga aatgaggata ttatttccag attgcaggag atggtagctg acctggagct
3121 gcagcaagac ctcatagtgc ctctcggcca caccccttcc caggagcact tcctctttga
3181 gattttccgc agacggctcc aggctctgac aagcgggtgg agcgtggctg ccagcctcta
3241 gagacagagg gagctgctaa tgtacaaacg gatcctcctc cgcctgcctt cgtctgtcct
3301 ctgcggcagc agcttccagg cagaacagcc catcactgcc agatgcgagc agttcttcca
3361 cttggtcaac tctgagatga gaaacttctg ctcccacgga ggtgccctga cacaggacat
3421 cactgcccac ttcttcaggg gcctcctgaa cgcctgtctg cggagcagag acccctccct
3481 gatggtcgac ttcatactgg ccaagtgcca gacgaaatgc cccttaattt tgacctctgc
3541 tctggtgtgg tggccgagcc tggagcctgt gctgctctgc cggtggagga gacactgcca
3601 gagcccgctg ccccgggaac tgcagaagct acaagaaggc cggcagtttg ccagcgattt
3661 cctctcccct gaggctgcct ccccagcacc caaccccgac tggctctcag ctgctgcact
3721 gcacttttgcg attcaacaag tcagggaaga aaacatcagg aagcagctaa agaagctgga
3781 ctgcgagaga gaggagctat tggttttcct tttcttcttc tccttgatgg gcctgctgtc
3841 gtcacatctg acctcaaata gcaccacaga cctgccaaag gctttccacg tttgtgcagc
3901 aatcctcgag tgtttagaga agaggaagat atcctggctg gcactctttc agttgacaga
3961 gagtgacctc aggctggggc ggctcctcct ccgtgtggcc ccggatcagc acaccaggct
4021 gctgcctttc gcttttaca gtcttctctc ctacttccat gaagacgcgg ccatcaggga
4081 agaggccttc ctgcatgttg ctgtggacat gtacttgaag ctggtccagc tcttcgtggc
4141 tggggataca agcacagttt cacctccagc tggcaggagc ctggagctca agggtcaggg
4201 caaccccgtg gaactgataa caaaagctcg tcttttttctg ctgcagttaa tacctcggtg
4261 cccgaaaaag agcttctcac acgtggcaga gctgctggct gatcgtgggg actgccgacc
4321 agaggtgagc gccgccctcc agagcagaca gcaggctgcc cctgacgctg acctgtccca
4381 ggagcctcat ctcttctgac gggcactgcc acgcccagct agcccagctc ccgtgtaaat
4441 aatttattac aagcataaca tggagctctt gttgcactaa aaagtggatt acaaatctcc
4501 tcgactgctt tagtggggaa aggaatcaat tatttatgaa ctgtccggcc ccgagtcact
4561 cagcgtttgc gggaaaataa accactggtc ccagagcaga ggaaggctac ttgagccgga
4621 caccaagccc ggctccagca ccaagggcgg gcagcaccct ccgaccctcc catgcgggtg
4681 cacacgaagg gtgaggctga cacagccact gcggagtcca ggctgctaga ggtgctcatc
4741 ctcactgccg tcctcaggtg ggttcgggct tcaccgcctg gccctctgtg gtcacagagg
4801 ggctcggtgg cccaggtggt ggttccgcct ccaggggcag ggccttgtcc tgggtctgtg
4861 tcagcgggtg caccatggac atgtgtacat tgaggttgtg ggccttctca aaccgccggc
4921 cacactggtc acaggcaaag tccagctcag tctcagcctt gtgtttggtc atgtggtact
4981 tgagggatgc ccgctgcctg cactggaacc cacagacctc acacctgggg gacagaggca
5041 gataagaagg tgcgaggcca cagcctctggg aggggggtcct gactcacact tactgcaaag
5101 gcttggctcc cgaatgtcgc atttggtgga cgagaaggtg cttccgctgc ttgaaggttt
5161 gtccacattc gtcacagata tagttccgca cctctgagag gggagagtcc agtgagtcca
5221 ggccctgat gctccaacct ccccgggga cgacgatgac aatgtgaaac catcacagct
5281 gggaagacat ttctgcacat ggttcaccat gcagtgggcc caagcaaggg gcctatgagg
5341 gcctcgttta ttaagatctt taaactgctt tatacactgt cacgtggctt catcagctgt
```

FIG. 7 (5/5)

```
5401 gtgcatttca ggatggtttt taaagaaacc tcagaaagct atttccttaa aaaaaaaaaa
5461 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaa
//
```

Revised: October 24, 2001.

FIG. 8 (1/4)



1: NM_030588. Homo sapiens DEAD...[gi:13514821]

DDX9

```
LOCUS       NM_030588               1378 bp    mRNA    linear   PRI 02-APR-2001
DEFINITION  Homo sapiens DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 9 (RNA
            helicase A, nuclear DNA helicase II; leukophysin) (DDX9),
            transcript variant 2, mRNA.
ACCESSION   NM_030588
VERSION     NM_030588.1  GI:13514821
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1378)
  AUTHORS   Lee,C.G. and Hurwitz,J.
  TITLE     A new RNA helicase isolated from HeLa cells that catalytically
            translocates in the 3' to 5' direction
  JOURNAL   J. Biol. Chem. 267 (7), 4398-4407 (1992)
  MEDLINE   92165790
   PUBMED   1537828
REFERENCE   2  (bases 1 to 1378)
  AUTHORS   Lee,C.G., Zamore,P.D., Green,M.R. and Hurwitz,J.
  TITLE     RNA annealing activity is intrinsically associated with U2AF
  JOURNAL   J. Biol. Chem. 268 (18), 13472-13478 (1993)
  MEDLINE   93293869
   PUBMED   7685763
REFERENCE   3  (bases 1 to 1378)
  AUTHORS   Lee,C.G. and Hurwitz,J.
  TITLE     Human RNA helicase A is homologous to the maleless protein of
            Drosophila
  JOURNAL   J. Biol. Chem. 268 (22), 16822-16830 (1993)
  MEDLINE   93346440
   PUBMED   8344961
REFERENCE   4  (bases 1 to 1378)
  AUTHORS   Abdelhaleem,M.M., Hameed,S., Klassen,D. and Greenberg,A.H.
  TITLE     Leukophysin: an RNA helicase A-related molecule identified in
            cytotoxic T cell granules and vesicles
  JOURNAL   J. Immunol. 156 (6), 2026-2035 (1996)
  MEDLINE   96310937
   PUBMED   8690889
REFERENCE   5  (bases 1 to 1378)
  AUTHORS   Zhang,S. and Grosse,F.
  TITLE     Domain structure of human nuclear DNA helicase II (RNA helicase A)
  JOURNAL   J. Biol. Chem. 272 (17), 11407-11494 (1997)
  MEDLINE   97269062
   PUBMED   9111062
REFERENCE   6  (bases 1 to 1378)
  AUTHORS   Nakajima,T., Uchida,C., Anderson,S.F., Lee,C.G., Hurwitz,J.,
```

FIG. 8 (2/4)

```
            Parvin,J.D. and Montminy,M.
  TITLE     RNA helicase A mediates association of CBP with RNA polymerase II
  JOURNAL   Cell 90 (6), 1107-1112 (1997)
  MEDLINE   97462911
  PUBMED    9323138
REFERENCE   7  (bases 1 to 1378)
  AUTHORS   Lee,C.G., da Costa Soares,V., Newberger,C., Manova,K., Lacy,E. and
            Hurwitz,J.
  TITLE     RNA helicase A is essential for normal gastrulation
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (23), 13709-13713 (1998)
  MEDLINE   99030634
  PUBMED    9811865
REFERENCE   8  (bases 1 to 1378)
  AUTHORS   Lee,C.G., Eki,T., Okumura,K., Nogami,M., Soares,Vd., Murakami,Y.,
            Hanaoka,F. and Hurwitz,J.
  TITLE     The human RNA helicase A (DDX9) gene maps to the prostate cancer
            susceptibility locus at chromosome band 1q25 and its pseudogene
            (DDX9P) to 13q22, respectively
  JOURNAL   Somat. Cell Mol. Genet. 25 (1), 33-39 (1999)
  MEDLINE   20381755
  PUBMED    10925702
COMMENT     REVIEWED REFSEQ: This record has been curated by NCBI staff. The
            reference sequence was derived from U03643.1.
            Summary: DEAD box proteins, characterized by the conserved motif
            Asp-Glu-Ala-Asp (DEAD), are putative RNA helicases. They are
            implicated in a number of cellular processes involving alteration
            of RNA secondary structure such as translation initiation, nuclear
            and mitochondrial splicing, and ribosome and spliceosome assembly.
            Based on their distribution patterns, some members of this family
            are believed to be involved in embryogenesis, spermatogenesis, and
            cellular growth and division. This gene includes 2 alternatively
            spliced transcripts, encoding 2 different isoforms. The larger
            isoform is a DEAD box protein with RNA helicase activity. It may
            participate in melting of DNA:RNA hybrids, such as those that occur
            during transcription, and may play a role in X-linked gene
            expression. It contains 2 copies of a double-stranded RNA-binding
            domain, a DEXH core domain and an RGG box. The RNA-binding domains
            and RGG box influence and regulate RNA helicase activity. The
            smaller isoform is a lymphocyte granule protein. It lacks
            RNA-binding domains and DEXH core domain, but contains an RGG box,
            which may render this isoform RNA binding function.
            Transcript Variant: This variant (2) is missing a 104 nt internal
            fragment, in addition to 2722 nt in the 5' UTR, as compared to
            variant 1. It encodes the smaller isoform, which is associated with
            lymphocyte granules.
            COMPLETENESS: complete on the 3' end.
FEATURES              Location/Qualifiers
     source           1..1378
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="1"
                      /map="1q25"
     gene             1..1378
                      /gene="DDX9"
                      /note="LKP; NDHII; RHA"
                      /db_xref="LocusID:1660"
                      /db_xref="MIM:603115"
     variation        35
                      /allele="A"
                      /allele="G"
```

FIG. 8 (3/4)

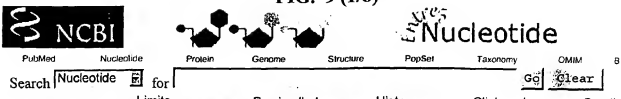
```
                     /db_xref="dbSNP:1049264"
     variation       51
                     /allele="A"
                     /allele="G"
                     /db_xref="dbSNP:1049265"
     variation       52
                     /allele="A"
                     /allele="G"
                     /db_xref="dbSNP:1049266"
     CDS             358..1065
                     /gene="DDX9"
                     /note="isoform 2 is encoded by transcript variant 2; RNA
                     helicase A; leukophysin; DEAD/H box-9; nuclear DNA
                     helicase II; ATP-dependent RNA helicase A"
                     /codon_start=1
                     /db_xref="LocusID:1660"
                     /db_xref="MIM:603115"
                     /product="DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 9,
                     isoform 2"
                     /protein_id="NP_085077.1"
                     /db_xref="GI:13514822"
                     /translation="MKYPSPFFVFGEKIRTRAISAKGMTLVTPLQLLLFASKKVQSDG
                     QIVLVDDWIKLQISHEAAACITGLRAAMEALVVEVTKQPAIISQLDPVNERMLNMIRQ
                     ISRPSAAGINLMIGSTRYGDGPRPPKMARYDNGSGYRRGGSSYSGGGYGGGYSSGGYG
                     SGGYGGSANSFRAGYGAGVGGGYRGVSRGGFRGNSGGDYRGPSGGYRGSGGFQRGGGR
                     GAYGTGYFGQGRGGGGY"
     misc_feature    760..1062
                     /note="Arg/Gly/Ser/Tyr-rich domain; Region: RGG box"
     variation       1146
                     /allele="A"
                     /allele="T"
                     /db_xref="dbSNP:861"
     variation       1187
                     /allele="G"
                     /allele="T"
                     /db_xref="dbSNP:865"
     variation       1236
                     /allele="G"
                     /allele="T"
                     /db_xref="dbSNP:860"
     variation       1240
                     /allele="A"
                     /allele="T"
                     /db_xref="dbSNP:864"
     variation       1293
                     /allele="C"
                     /allele="T"
                     /db_xref="dbSNP:863"
     variation       1297
                     /allele="A"
                     /allele="T"
                     /db_xref="dbSNP:866"
     variation       1318
                     /allele="A"
                     /allele="T"
                     /db_xref="dbSNP:862"
     polyA_signal    1362..1367
     polyA_site      1378
BASE COUNT      369 a    261 c    351 g    397 t
ORIGIN
```

FIG. 8 (4/4)

```
   1 cattgctgct gctacctgct ttccagagcc tttcatcaat gaaggaaagc ggctgggcta
  61 tatccatcga aattttgctg gaaacagatt ttctgatcac gtagcccttt tatcagtatt
 121 ccaagcctgg gatgatgcta gaatgggtgg agaagaagca gagatacgtt tttgtgagca
 181 caaaagactt aatatggcta cactaagaat gacctgggaa gccaaagttc agctcaaaga
 241 gattttgatt aattctgggt ttccagaaga ttgtttgttg acacaagtgt ttactaacac
 301 tggaccagat aataatttgg atgttgttat ctccctcctg gcctttgtag ccaagacatg
 361 aagtacccat ctcccttctt tgtatttggt gaaaagattc gaactcgagc catctctgct
 421 aaaggcatga ctttagtcac cccctgcag ttgcttctct ttgcctccaa gaaagtccaa
 481 tctgatgggc agattgtgct tgtagatgac tggattaaac tgcaaatatc tcatgaagct
 541 gctgcctgta tcactggtct ccgggcagcc atggaggctt tggttgttga agtaaccaaa
 601 caacctgcta tcatcagcca gttggacccc gtaaatgaac gtatgctgaa catgatccgt
 661 cagatctcta gaccctcagc tgctggtatc aaccttatga ttggcagtac acggtatgga
 721 gatggtccac gtcctcccaa gatggcccga tacgacaatg gaagcggata tagaagggga
 781 ggttctagtt acagtggtgg aggctatggc ggtggctata gcagtggagg ctatggtagc
 841 ggaggctatg gtggcagcgc caactccttt cgggcaggat atggtgcagg tgttggtgga
 901 ggctatagag gagtttcccg aggtggcttt agaggcaact ctggaggaga ctacagaggg
 961 cctagtggag gctacagagg atctgggggа ttccagcgag gaggtggtag gggggcctat
1021 ggaactggct actttggaca gggaagagga ggtggcggct attaaaactt ggttatgtca
1081 gttcctgtgt gtagacagta aggaaaaaaa ggcatgctat gtgttacgtg ttttttccag
1141 tatgtttatt tgccaccaaa aagtaaatgc attttcaccc attctgtggt tcattgtagt
1201 ttaaggaaac caagcatata gatgcattag tgatttttgtt tatattattgt aaaatataac
1261 gatcttctaa aaataccaca gtttgtattt tttctttaag gagtaaagat ttgcctttaa
1321 ataacttggt attttcctgg ctttcgttta atacaataga aaataaagta ttacaccg
//
```

Revised: October 24, 2001.

FIG. 9 (1/6)



1: NM_000875. Homo sapiens insu...[gi:11068002]

Related Sequences, OMIM, Protein, PubMed, Taxonomy, UniSTS, LinkOut

IGFI-R

```
LOCUS       NM_000875                4989 bp    mRNA    linear   PRI 01-NOV-2000
DEFINITION  Homo sapiens insulin-like growth factor 1 receptor (IGF1R), mRNA.
ACCESSION   NM_000875
VERSION     NM_000875.2  GI:11068002
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4989)
  AUTHORS   Flier JS, Usher P and Moses AC.
  TITLE     Monoclonal antibody to the type I insulin-like growth factor
            (IGF-I) receptor blocks IGF-I receptor-mediated DNA synthesis:
            clarification of the mitogenic mechanisms of IGF-I and insulin in
            human skin fibroblasts
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 83 (3), 664-668 (1986)
  MEDLINE   86121000
   PUBMED   3003744
REFERENCE   2  (bases 1 to 4989)
  AUTHORS   Francke U, Yang-Feng TL, Brissenden JE and Ullrich A.
  TITLE     Chromosomal mapping of genes involved in growth control
  JOURNAL   Cold Spring Harb. Symp. Quant. Biol. 51 Pt 2, 855-866 (1986)
  MEDLINE   87217109
   PUBMED   3107886
REFERENCE   3  (bases 1 to 4989)
  AUTHORS   Ullrich,A., Gray,A., Tam,A.W., Yang-Feng,T., Tsubokawa,M.,
            Collins,C., Henzel,W., Bon,T.L., Kathuria,S., Chen,E., Jakobs,S.,
            Francke,U., Ramachandran,J. and Fujita-Yamaguchi,Y.
  TITLE     Insulin-like growth factor I receptor primary structure: comparison
            with insulin receptor suggests structural determinants that define
            functional specificity
  JOURNAL   EMBO J. 5 (10), 2503-2512 (1986)
  MEDLINE   87053815
REFERENCE   4  (bases 1 to 4989)
  AUTHORS   Cooke DW, Bankert LA, Roberts CT Jr, LeRoith D and Casella SJ.
  TITLE     Analysis of the human type I insulin-like growth factor receptor
            promoter region
  JOURNAL   Biochem. Biophys. Res. Commun. 177 (3), 1113-1120 (1991)
  MEDLINE   91282751
   PUBMED   1711844
REFERENCE   5  (bases 1 to 4989)
  AUTHORS   Abbott AM, Bueno R, Pedrini MT, Murray JM and Smith RJ.
  TITLE     Insulin-like growth factor I receptor gene structure
  JOURNAL   J. Biol. Chem. 267 (15), 10759-10763 (1992)
  MEDLINE   92268129
   PUBMED   1316909
```

FIG. 9 (2/6)

```
REFERENCE   6  (bases 1 to 4989)
  AUTHORS   Werner H, Karnieli E, Rauscher FJ and LeRoith D.
  TITLE     Wild-type and mutant p53 differentially regulate transcription of
            the insulin-like growth factor I receptor gene
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 93 (16), 8318-8323 (1996)
  MEDLINE   96323219
   PUBMED   8710868
REFERENCE   7  (bases 1 to 4989)
  AUTHORS   Grant ES, Ross MB, Ballard S, Naylor A and Habib FK.
  TITLE     The insulin-like growth factor type I receptor stimulates growth
            and suppresses apoptosis in prostatic stromal cells
  JOURNAL   J. Clin. Endocrinol. Metab. 83 (9), 3252-3257 (1998)
  MEDLINE   98417960
   PUBMED   9745438
COMMENT     REVIEWED REFSEQ: This record has been curated by NCBI staff. The
            reference sequence was derived from X04434.1, M69229.1.
            On Nov 1, 2000 this sequence version replaced gi:4557664.
            Summary: This receptor binds insulin-like growth factor with a high
            affinity. It has tyrosine kinase activity. The insulin-like growth
            factor I receptor plays a critical role in transformation events.
            Cleavage of the precursor generates alpha and beta subunits. It is
            highly overexpressed in most malignant tissues where it functions
            as an anti-apoptotic agent by enhancing cell survival.
FEATURES             Location/Qualifiers
     source          1..4989
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15q25-q26"
                     /clone="(lambda)IGF-1-R.85, (lambda)IGF-1-R.76"
                     /tissue_type="placenta"
                     /clone_lib="(lamda)gt10"
     gene            1..4989
                     /gene="IGF1R"
                     /note="JTK13"
                     /db_xref="LocusID:3480"
                     /db_xref="MIM:147370"
     CDS             46..4149
                     /gene="IGF1R"
                     /EC_number="2.7.1.112"
                     /codon_start=1
                     /db_xref="LocusID:3480"
                     /db_xref="MIM:147370"
                     /product="insulin-like growth factor 1 receptor precursor"
                     /protein_id="NP_000866.1"
                     /db_xref="GI:4557665"
                     /translation="MKSGSGGGSPTSLWGLLFLSAALSLWPTSGEICGPGIDIRNDYQ
                     QLKRLENCTVIEGYLHILLISKAEDYRSYRFPKLTVITEYLLLFRVAGLESLGDLFPN
                     LTVIRGWKLFYNYALVIFEMTNLKDIGLYNLRNITRGAIRIEKNADLCYLSTVDWSLI
                     LDAVSNNYIVGNKPPKECGDLCPGTMEEKPMCEKTTINNEYNYRCWTTNRCQKMCPST
                     CGKRACTENNECCHPECLGSCSAPDNDTACVACRHYYYAGVCVPACPPNTYRFEGWRC
                     VDRDFCANILSAESSDSEGFVIHDGECMQECPSGFIRNGSQSMYCIPCEGPCPKVCEE
                     EKKTKTIDSVTSAQMLQGCTIFKGNLLINIRRGNNIASELENFMGLIEVVTGYVKIRH
                     SHALVSLSFLKNLRLILGEEQLEGNYSFYVLDNQNLQQLWDWDHRNLTIKAGKMYFAF
                     NPKLCVSEIYRMEEVTGTKGRQSKGDINTRNNGERASCESDVLHFTSTTTSKNRIIIT
                     WHRYRPPDYRDLISFTVYYKEAPFKNVTEYDGQDACGSNSWNMVDVDLPPNKDVEPGI
                     LLHGLKPWTQYAVYVKAVTLTMVENDHIRGAKSEILYIRTNASVPSIPLDVLSASNSS
                     SQLIVKWNPPSLPNGNLSYYIVRWQRQPQDGYLYRHNYCSKDKIPIRKYADGTIDIEE
                     VTENPKTEVCGGEKGPCCACPKTEAEKQAEKEEAEYRKVFENFLHNSIFVPRPERKRR
                     DVMQVANTTMSSRSRNTTAADTYNITDPEELETEYPFFESRVDNKERTVISNLRPFTL
```

FIG. 9 (3/6)

YRIDIHSCNHEAEKLGCSASNFVFARTMPAEGADDIPGPVTWEPRPENSIFLKWPEPE
NPNGLILMYEIKYGSQVEDQRECVSRQEYRKYGGAKLNRLNPGNYTARIQATSLSGNG
SWTDPVFFYVQAKTGYENFIHLIIALPVAVLLIVGGLVIMLYVFHRKRNNSRLGNGVL
YASVNPEYFSAADVYVPDEWEVAREKITMSRELGQGSFGMVYEGVAKGVVKDEPETRV
AIKTVNEAASMRERIEFLNEASVMKEFNCHHVVRLLGVVSQGQPTLVIMELMTRGDLK
SYLRSLRPEMENNPVLAPPSLSKMIQMAGEIADGMAYLNANKFVHRDLAARNCMVAED
FTVKIGDFGMTRDIYETDYYRKGGKGLLPVRWMSPESLKDGVFTTYSDVWSFGVVLWE
IATLAEQPYQGLSNEQVLRFVMEGGLLDKPDNCPDMLFELMRMCWQYNPKMRPSFLEI
ISSIKEEMEPGFREVSFYYSEENKLPEPEELDLEPENMESVPLDPSASSSSLPLPDRH
SGHKAENGPGPGVLVLRASFDERQPYAHMNGGRKNERALPLPQSSTC"

sig_peptide 46..123
mat_peptide 121..4134
/product="IGF-I receptor"
misc_feature 122..2251
/note="alpha-subunit (AA 1 - 710)"
misc_feature 182..190
/note="pot.N-linked glycosylation site (AA 21 - 23)"
misc_feature 196..561
/note="Recep_L_domain; Region: Receptor L domain"
misc_feature 335..343
/note="pot.N-linked glycostlation site (AA 72 - 74)"
misc_feature 434..442
/note="pot.N-linked glycostlation site (AA 105 - 107)"
misc_feature 568..1044
/note="Furin-like; Region: Furin-like cysteine rich region"
misc_feature 724..852
/note="FU; Region: Furin-like repeats"
misc_feature 761..769
/note="pot.N-linked glycostlation site (AA 214 - 216)"
misc_feature 971..979
/note="pot.N-linked glycostlation site (AA 284 - 286)"
misc_feature 1162..1410
/note="Recep_L_domain; Region: Receptor L domain"
misc_feature 1280..1288
/note="pot.N-linked glycostlation site (AA 387 - 389)"
misc_feature 1343..1351
/note="pot.N-linked glycosylation site (AA 408 - 410)"
misc_feature 1631..1639
/note="pot.N-linked glycostlation site (AA 504 - 506)"
variation 1731
/allele="A"
/allele="G"
/db_xref="dbSNP:2228531"
misc_feature 1850..1858
/note="pot.N-linked glycosylation site (AA 577 - 579)"
misc_feature 1895..1903
/note="pot.N-linked glycosylation site (AA 592 - 594)"
misc_feature 1949..1957
/note="pot.N-linked glycosylation site (AA 610 - 612)"
misc_feature 2240..2251
/note="putative proreceptor processing site (AA 707 - 710)"
misc_feature 2252..4132
/note="beta-subunit (AA 711 - 1337)"
misc_feature 2270..2278
/note="pot.N-linked glycosylation site (AA 717 - 719]"
misc_feature 2297..2305
/note="pot.N-linked glycosylation site (AA 726 - 728)"
misc_feature 2321..2329

FIG. 9 (4/6)

```
                     /note="pot.N-linked glycosylation site (AA 734 - 736)
     misc_feature    2548..2796
                     /note="fn3; Region: Fibronectin type III domain"
     misc_feature    2729..2737
                     /note="pot.N-linked glycosylation site (AA 870 - 872)"
     misc_feature    2768..2776
                     /note="pot.N-linked glycosylation site (AA 883 - 885)"
     misc_feature    2836..2910
                     /note="transmembrane region (AA 906 - 929);
                     transmembrane-region site"
     misc_feature    2918..2926
                     /note="pot.N-linked glycosylation site (AA 933 - 935)"
     misc_feature    3040..3834
                     /note="pkinase; Region: Eukaryotic protein kinase domain"
     misc_feature    3040..3843
                     /note="TyrKc; Region: Tyrosine kinase, catalytic domain"
     misc_feature    3047..3049
                     /note="pot.ATP binding site (AA 976)"
     misc_feature    3052..3807
                     /note="S_TKc; Region: Serine/Threonine protein kinases,
                     catalytic domain"
     misc_feature    3053..3055
                     /note="pot.ATP binding site (AA 978)"
     misc_feature    3062..3064
                     /note="pot.ATP binding site (AA 981)"
     misc_feature    3128..3130
                     /note="pot.ATP binding site (AA 1003)"
     variation       4267
                     /allele="A"
                     /allele="T"
                     /db_xref="dbSNP:1065304"
     variation       4268
                     /allele="A"
                     /allele="T"
                     /db_xref="dbSNP:1065305"
BASE COUNT      1216 a    1371 c    1320 g    1082 t
ORIGIN
        1 ttttttttt ttttgagaaa gggaatttca tcccaaataa aaggaatgaa gtctggctcc
       61 ggaggagggt ccccgacctc gctgtggggg ctcctgtttc tctccgccgc gctctcgctc
      121 tggccgacga gtggagaaat ctgcgggcca ggcatcgaca tccgcaacga ctatcagcag
      181 ctgaagcgcc tggagaactg cacggtgatc gagggctacc tccacatcct gctcatctcc
      241 aaggccgagg actaccgcag ctaccgcttc cccaagctca cggtcattac cgagtacttg
      301 ctgctgttcc gagtggctgg cctcgagagc ctcggagacc tcttccccaa cctcacggtc
      361 atccgcggct ggaaactctt ctacaactac gccctggtca tcttcgagat gaccaatctc
      421 aaggatattg ggctttacaa cctgaggaac attactcggg gggccatcag gattgagaaa
      481 aatgctgacc tctgttacct ctccactgtg gactggtccc tgatcctgga tgcggtgtcc
      541 aataactaca ttgtggggaa taagccccca aaggaatgtg gggacctgtg tccagggacc
      601 atggaggaga agccgatgtg tgagaagacc accatcaaca atgagtacaa ctaccgctgc
      661 tggaccacaa accgctgcca gaaaatgtgc ccaagcacgt gtgggaagcg ggcgtgcacc
      721 gagaacaatg agtgctgcca ccccgagtgc ctgggcagct gcagcgcgcc tgacaacgac
      781 acggcctgtg tagcttgccg ccactactac tatgccggtg tctgtgtgcc tgcctgcccg
      841 cccaacacct acaggtttga gggctggcgc tgtgtggacc gtgacttctg cgccaacatc
      901 ctcagcgccg agagcagcga ctccgagggg tttgtgatcc acgacggcga gtgcatgcag
      961 gagtgcccct cgggcttcat ccgcaacggc agccagagca tgtactgcat cccttgtgaa
     1021 ggtcctgcc cgaaggtctg tgaggaagaa aagaaaacaa agaccattga ttctgttact
     1081 tctgctcaga tgctccaagg atgcaccatc ttcaagggca atttgctcat taacatccga
     1141 cgggggaata acattgcttc agagctggag aacttcatgg ggctcatcga ggtggtgacg
     1201 ggctacgtga agatccgcca ttctcatgcc ttggtctcct tgtccttcct aaaaaacctt
     1261 cgcctcatcc taggagagga gcagctagaa gggaattact ccttctacgt cctcgacaac
     1321 cagaacttgc agcaactgtg ggactgggac caccgcaacc tgaccatcaa agcagggaaa
```

FIG. 9 (5/6)

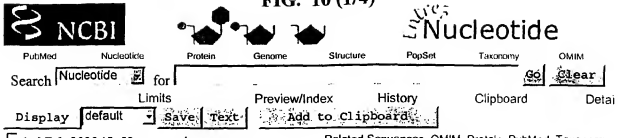
```
1381 atgtactttg ctttcaatcc caaattatgt gtttccgaaa tttaccgcat ggaggaagtg
1441 acgggggacta aagggcgcca aagcaaaggg gacataaaca ccaggaacaa cggggagaga
1501 gcctcctgtg aaagtgacgt cctgcatttc acctccacca ccacgtcgaa gaatcgcatc
1561 atcataacct ggcaccggta ccggccccct gactacaggg atctcatcag cttcaccgtt
1621 tactacaagg aagcaccctt taagaatgtc acagagtatg atgggcagga tgcctgcggc
1681 tccaacagct ggaacatggt ggacgtggac ctcccgccca acaaggacgt ggagccctgc
1741 atcttactac atgggctgaa gccctggact cagtacgccg tttacgtcaa ggctgtgacc
1801 ctcaccatgg tggagaacga ccatatccgt ggggccaaga gtgagatctt gtacattcgc
1861 accaatgctt cagttccttc cattcccttg gacgttcttt cagcatcgaa ctcctcttct
1921 cagttaatcg tgaagtggaa ccctcctct ctgcccaacg gcaacctgag ttactacatt
1981 gtgcgctggc agcggcagcc tcaggacggc taccttaacc ggcacaatta ctgctccaaa
2041 gacaaaatcc ccatcaggaa gtatgccgac ggcaccatcg acattgagga ggtcacagag
2101 aaccccaaga ctgaggtgtg tggtggggag aaagggcctt gctgcgcctg ccccaaaact
2161 gaagccgaga agcaggccga gaaggaggag gctgaatacc gcaaagtctt tgagaatttc
2221 ctgcacaact ccatcttcgt gcccagacct gaaaggaagc ggagagatgt catgcaagtg
2281 gccaacacca ccatgtccag ccgaagcagg aacaccacgg ccgcagacac ctacaacatc
2341 accgacccgg aagagctgga gacagagtac cctttctttg agagcagagt ggataacaag
2401 gagagaactg tcatttctaa ccttcggcct ttcacattgt accgcatcga tatccacagc
2461 tgcaaccacg aggctgagaa gctgggctgc agcgcctcca acttcgtctt tgcaaggact
2521 atgcccgcag aaggagcaga tgacattcct gggccagtga cctgggagcc aaggcctgaa
2581 aactccatct ttttaaagtg gcggaacct gagaatccca atggattgat tctaatgtat
2641 gaaataaaat acggatcaca agttgaggat cagcgagaat gtgtgtccag acaggaatac
2701 aggaagtatg gaggggccaa gctaaaccgg ctaaacccgg ggaactacac agcccggatt
2761 caggccacat ctctctctgg gaatgggtcg tggacagatc ctgtgttctt ctatgtccag
2821 gccaaaacag gatatgaaaa cttcatccat ctgatcatcg ctctgcccgt cgctgtcctg
2881 ttgatcgtgg gagggttggt gattatgctg tacgtcttcc atagaaagag aaataacagc
2941 aggctgggga atggagtgct gtatgcctct gtgaacccgg agtacttcag cgctgctgat
3001 gtgtacgttc ctgatgagtg ggaggtggct cgggagaaga tcaccatgag ccgggaactt
3061 gggcaggggt cgtttgggat ggtctatgaa ggagttgcca agggtgtggt gaaagatgaa
3121 cctgaaacca gagtggccat taaaacagtg aacgaggccg caagcatgcg tgagaggatt
3181 gagtttctca acgaagcttc tgtgatgaag gagttcaatt gtcaccatgt ggtgcgattg
3241 ctgggtgtgg tgtcccaagg ccagccaaca ctggtcatca tggaactgat gacacggggc
3301 gatctcaaaa gttatctccg gtctctgagg ccagaaatgg agaatcatcc agtcctagca
3361 cctccaagcc tgagcaagat gattcagatg gccggagaga ttgcagacgg catggccatac
3421 ctcaacgcca ataagttcgt ccacagagac cttgctgccc ggaattgcat ggtagccgaa
3481 gatttcacag tcaaaatcgg agattttgat gtgacgcgag atatctatga gacagactat
3541 taccggaaag gaggcaaagg gctgctgccc gtgcgctgga tgtctcctga gtccctcaag
3601 gatggagtct tcaccactta ctggacgtc tggtccttcg gggtcgtcct ctgggagatc
3661 gccacactgg ccgagcagcc ctaccagggc ttgtccaacg agcaagtcct tcgcttcgtc
3721 atggagggcg gccttctgga caagccagac aactgtcctg acatgctgtt tgaactgatg
3781 cgcatgtgct ggcagtataa ccccaagatg aggccttcct tcctggagat catcagcagc
3841 atcaaagagg agatggagcc tggcttccgg gaggtctcct tctactacag cgaggagaac
3901 aagctgcccg agccggagga gctggacctg gagccagaga acatggagag cgtcccctg
3961 gaccccctgg cctcctcgtc ctccctgcca ctgccccaca gacactcagg acacaaggcc
4021 gagaacggcc ccggccctgg ggtgctggtc ctccgcgcca gcttcgacga gagacagcct
4081 tacgcccaca tgaacggggg ccgcaagaac gagcgggcct tgccgctgcc ccagtcttcg
4141 acctgctgat ccttggatcc tgaatctgtg caaacagtaa cgtgtgcgca cgcgcagcgg
4201 ggtgggggggg gagagagagt tttaacaatc cattcacaag cctcctgtac ctcagtggat
4261 cttcagttct gcccttgctg cccgcgggag acagcttctc tgcagtaaaa cacatttggg
4321 atgttccttt tttcaatatg caagcagctt tttattccct gcccaaaccc ttaactgaca
4381 tgggccttta agaaccttaa tgacaacact taatagcaac agagcacttg agaaccagtc
4441 tcctcactct gtccctgtcc ttccctgttc tcccttctc tctcctctct gcttcataac
4501 ggaaaaatac ttgccacaag tccagctggg aagccctttt tatcagtttg aggaagtggc
4561 tgtccctgtg gcccatcca accactgtac acacccgcct gacaccgtgg gtcattacaa
4621 aaaaacacgt ggagatggaa atttttacct ttatctttca cctttctagg gacatgaaat
4681 ttacaaaggg ccatcgttca tccaaggctg ttaccatttt aacgctgcct aattttgcca
4741 aaatcctgaa ctttctccct catcggcccg gcgctgattc ctcgtgtccg gaggcatggg
4801 tgagcatggc agctggttgc tccatttgag agacacgctg gcgacacact ccgtccatcc
4861 gactgccccc gctgtgctgc tcaaggccac aggcacacag gtctcattgc ttctgactag
4921 attattattt gggggaactg gacacaatag gtctttctct cagtgaaggt ggggagaagc
```

FIG. 9 (6/6)

```
4981 tgaaccggc
//
```

Revised: October 24, 2001.

FIG. 10 (1/4)



LOCUS NM_003349 2394 bp mRNA linear PRI 21-SEP-2001
DEFINITION Homo sapiens ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1), transcript variant 2, mRNA.
ACCESSION NM_003349
VERSION NM_003349.3 GI:15718757

UBE2 V1

KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2394)
AUTHORS Rothofsky,M.L. and Lin,S.L.
TITLE CROC-1 encodes a protein which mediates transcriptional activation of the human FOS promoter
JOURNAL Gene 195 (2), 141-149 (1997)
MEDLINE 97449289
PUBMED 9305758
REFERENCE 2 (bases 1 to 2394)
AUTHORS Sancho, E., Vila, M.R., Sanchez-Pulido, L., Lozano, J.J., Paciucci, R., Nadal, M., Fox, M., Harvey, C., Bercovich, B., Loukili, N., Ciechanover, A., Lin, S.L., Sanz, F., Estivill, X., Valencia, A. and Thomson, T.M.
TITLE Role of UEV-1, an inactive variant of the E2 ubiquitin-conjugating enzymes, in in vitro differentiation and cell cycle behavior of HT-29-M6 intestinal mucosecretory cells
JOURNAL Mol. Cell. Biol. 18 (1), 576-589 (1998)
MEDLINE 98078713
PUBMED 9418904
REFERENCE 3 (bases 1 to 2394)
AUTHORS Ma,L., Broomfield,S., Lavery,C., Lin,S.L., Xiao,W. and Bacchetti,S.
TITLE Up-regulation of CIR1/CROC1 expression upon cell immortalization and in tumor-derived human cell lines
JOURNAL Oncogene 17 (10), 1321-1326 (1998)
MEDLINE 98442973
PUBMED 9771976
REFERENCE 4 (bases 1 to 2394)
AUTHORS Hofmann,R.M. and Pickart,C.M.
TITLE Noncanonical MMS2-encoded ubiquitin-conjugating enzyme functions in assembly of novel polyubiquitin chains for DNA repair
JOURNAL Cell 96 (5), 645-653 (1999)
MEDLINE 99189750
PUBMED 10089880
REFERENCE 5 (bases 1 to 2394)
AUTHORS Deng,L., Wang,C., Spencer,E., Yang,L., Braun,A., You,J., Slaughter,C., Pickart,C. and Chen,Z.J.
TITLE Activation of the IkappaB kinase complex by TRAF6 requires a

FIG. 10 (2/4)

dimeric ubiquitin-conjugating enzyme complex and a unique polyubiquitin chain

JOURNAL Cell 103 (2), 351-361 (2000)
MEDLINE 20509589
PUBMED 11057907
REFERENCE 6 (bases 1 to 2394)
AUTHORS Thomson,T.M., Lozano,J.J., Loukili,N., Carrio,R., Serras,F., Cormand,B., Valeri,M., Diaz,V.M., Abril,J., Burset,M., Merino,J., Macaya,A., Corominas,M. and Guigo,R.
TITLE Fusion of the human gene for the polyubiquitination coeffector UEV1 with Kua, a newly identified gene
JOURNAL Genome Res. 10 (11), 1743-1756 (2000)
MEDLINE 20530912
PUBMED 11076860
COMMENT REVIEWED REFSEQ: This record has been curated by NCBI staff. The reference sequence was derived from U39361.1, AL110132.1.
On Sep 21, 2001 this sequence version replaced gi:12025659.
Summary: Ubiquitin-conjugating enzyme E2 variant proteins constitute a distinct subfamily within the E2 protein family. They have sequence similarity to other ubiquitin-conjugating enzymes but lack the conserved cysteine residue that is critical for the catalytic activity of E2s. The protein encoded by this gene is located in the nucleus and can cause transcriptional activation of the human FOS proto-oncogene. It is thought to be involved in the control of differentiation by altering cell cycle behaviour. Multiple alternatively spliced transcripts encoding different isoforms have been described for this gene.
Transcript Variant: This variant (2) encodes the longest isoform (b) of this protein.
COMPLETENESS: complete on the 3' end.

FEATURES Location/Qualifiers
source 1..2394
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.2"
gene 1..2394
/gene="UBE2V1"
/note="CIR1; UEV-1; UEV1; UEV1A; CROC-1; CROC1"
/db_xref="LocusID:7335"
/db_xref="MIM:602995"
CDS 70..735
/gene="UBE2V1"
/note="isoform b is encoded by transcript variant 2; DNA-binding protein"
/codon_start=1
/db_xref="LocusID:7335"
/db_xref="MIM:602995"
/product="ubiquitin-conjugating enzyme E2 variant 1, isoform b"
/protein_id="NP_003340.1"
/db_xref="GI:4507795"
/translation="MAYKFRTHSPEALEQLYPWECPVFCLIIFGTFTNQIHKWSHTYF
GLPRWVTLLQDWHVILPRKHHRIHHVSPHETYFCITTGVKVPRNFRLLEELEEGQKGV
GDGTVSWGLEDDEDMTLTRWTGMIIGPPRTIYENRIYSLKIECGPKYPEAPPFVRFVT
KINMNGVNSSNGVVDPRAISVLAKWQNSYSIKVVLQELRRLMMSKENMKLPQPPEGQC
YSN"
misc_feature 334..714
/note="UBCc; Region: Ubiquitin-conjugating enzyme E2, catalytic domain homologues"

FIG. 10 (3/4)

```
     misc_feature    337..555
                     /note="UQ_con; Region: Ubiquitin-conjugating enzyme.
                     Proteins destined for proteasome-mediated degradation may
                     be ubiquitinated. Ubiquitination follows conjugation of
                     ubiquitin to a conserved cysteine residue of UBC
                     homologues. TSG101 is one of several UBC homologues that
                     lacks this active site cysteine"
     misc_feature    643..714
                     /note="Region: DNA-binding domain"
     variation       1117
                     /allele="C"
                     /allele="T"
                     /db_xref="dbSNP:8585"
     variation       1257
                     /allele="C"
                     /allele="T"
                     /db_xref="dbSNP:1049679"
     variation       complement(1968)
                     /allele="A"
                     /allele="G"
                     /db_xref="dbSNP:2733"
     variation       2017
                     /allele="A"
                     /allele="C"
                     /db_xref="dbSNP:15218"
     polyA_signal    2112..2117
     polyA_site      2135
                     /evidence=experimental
     variation       complement(2179)
                     /allele="G"
                     /allele="T"
                     /db_xref="dbSNP:2664563"
     variation       2249
                     /note="WARNING: map location ambiguous"
                     /allele="A"
                     /allele="T"
                     /db_xref="dbSNP:1049871"
     variation       complement(2259)
                     /allele="A"
                     /allele="G"
                     /db_xref="dbSNP:2664532"
     polyA_signal    2350..2355
                     /evidence=experimental
     polyA_site      2373
BASE COUNT      658 a    605 c    481 g    650 t
ORIGIN
        1 ttcacacggc acgacttcat cgagaccaac ggggacaact gcctggtgac actgctgccg
       61 ctgctaaaca tggcctacaa gttccgcacc cacagccctg aagccctgga gcagctatac
      121 ccctgggagt gcttcgtctt ctgcctgatc atcttcggca ccttcaccaa ccagatccac
      181 aagtggtcgc acacgtactt tgggctgcca cgctggctca ccctcctgca ggactggcat
      241 gtcatcctgc cacgtaaaca ccatcggatc caccacgtct caccccacga gacctacttc
      301 tgcatcacca caggagtaaa agtccctcgc aatttccgac tgttggaaga actcgaagaa
      361 ggccagaaag gagtaggaga tggcacagtt agctggggtc tagaagatga cgaagacatg
      421 acacttacaa gatggacagg gatgataatt gggcctccaa gaacaattta tgaaaaccga
      481 atatacagcc ttaaaataga atgtggacct aaatacccag aagcaccccc ctttgtaaga
      541 tttgtaacaa aaattaatat gaatggagta aatagttcta atggagtggt ggacccaaga
      601 gccatatcag tgctagcaaa atggcagaat tcatatagca tcaaagttgt cctgcaagag
      661 cttcggcgcc taatgatgtc taaagaaaat atgaaactcc ctcagccgcc cgaaggacag
      721 tgttacagca attaatcaaa aagaaaaacc acaggccctt ccccttcccc ccaattcgat
      781 ttaatcagtc ttcattttcc acagtagtaa attttctaga tacgtcttgt agacctcaaa
```

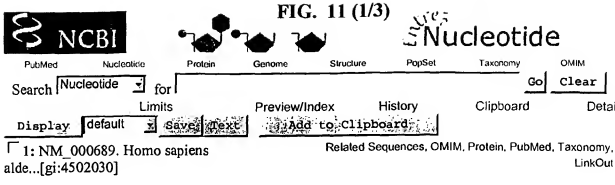
FIG. 10 (4/4)

```
 841 gtaccggaaa ggaagctccc attcaaagga aatttatctt aagatactgt aaatgatact
 901 aatttttttgt ccatttgaaa tatataagtt gtgctataac aaatcatcct gtcaagtgta
 961 accactgtcc acgtagttga acttctggga tcaagaaagt ctatttaaat tgattcccat
1021 cataactggt ggggcacatc taactcaact gtgaaaagac acatcacaca atcaccttgc
1081 tgctgattac acggcctggg gtctctgcct tctcccttta ccctcccgcc tcccaccctc
1141 cctgcaacaa cagccctcta gcctgggggg cttgttagag tagatgtgaa ggtttcaggt
1201 cgcagcctgt gggactactg ctaggtgtgt ggggtgtttc gcctgcaccc ctggttcctt
1261 taagtcttaa gtgatgcccc ttccaaacca tcatcctgtc cccacgctcc tccactcccg
1321 cccttggccg aagcatagat tgtaacccct ccactcccct ctgagattgg cttcggtgag
1381 gaattcaggg ctttccccat atcttctctc cccccacctt tatcgagggg tgctgctttt
1441 tctccctcct cctcaagttc ctttttgcac cgtcaccacc caacaccttc catgacactt
1501 ccttgctttg gccagaagcc atcaggtaag gttggaaaga gcctctgacc tcccttgttt
1561 agtttttggaa ccatactcac tcactctcca ccagcctggg aaatgaatat tgggtcctca
1621 gccctgccac cctctgctgt catcagctga tgcattgttt ttagctcagg ttttgataag
1681 gtgaaaagaa tagtcaccag ggttactcag acctgccagc tctcggagtc cttggtggtt
1741 gaacttggag aaagaccgca tgaagatact tgtaagcaca catgatccct ctgaattgtt
1801 ttactttcct gtaactgctt ttgcttttaa aaattgaaga agttttaaac agggctttca
1861 tttggtcatc cttgcaatcc attggggtct agtttggaat ctgacaactg gaacaaaaag
1921 aaccttgaat ccggtgcatg ccttggtttt ggtgctgctg ctgcttccca agatcctcag
1981 cagggattaa gaaggaaccc ggtgtgcaca gcagatcccc gaaattggtg ggcttgacct
2041 cctggcaaat tgctgcgtct ttccacttgc tgttcaggac cactaaatgc tgaaatgtgg
2101 atgcataccg aaataaaagc aattcattgt gtactaaagg ttttttttct ttttttaatt
2161 tagtatttgt gtaaaaccac cttttgaagc agcaactatc aagtctgaaa agcaattgat
2221 gtttccatta atctttttct gggggaaaa ccttagttct aaggatttaa catcctgtaa
2281 gtgaagttta acataacagt attccataag cagccttttt attgtcagac cattgcctga
2341 ttttaatata ataaaaaaaa agtgtgcgtt aataaaaaaa aaaaaaaaa aaaa
```

//

Revised: October 24, 2001.

FIG. 11 (1/3)



LOCUS NM_000689 1506 bp mRNA linear PRI 31-OCT-2000
DEFINITION Homo sapiens aldehyde dehydrogenase 1, soluble (ALDH1), mRNA.
ACCESSION NM_000689
VERSION NM_000689.1 GI:4502030
KEYWORDS .
SOURCE human.

ALDEHYDE DEHYDROGENASE

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1506)
AUTHORS Hsu LC, Tani K, Fujiyoshi T, Kurachi K and Yoshida A.
TITLE Cloning of cDNAs for human aldehyde dehydrogenases 1 and 2
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 82 (11), 3771-3775 (1985)
MEDLINE 85216574
PUBMED 2987944
REFERENCE 2 (bases 1 to 1506)
AUTHORS Raghunathan L, Hsu LC, Klisak I, Sparkes RS, Yoshida A and Mohandas T.
TITLE Regional localization of the human genes for aldehyde dehydrogenase-1 and aldehyde dehydrogenase-2
JOURNAL Genomics 2 (3), 267-269 (1988)
MEDLINE 88284707
PUBMED 3397064
REFERENCE 3 (bases 1 to 1506)
AUTHORS Hsu LC, Chang WC and Yoshida A.
TITLE Genomic structure of the human cytosolic aldehyde dehydrogenase gene
JOURNAL Genomics 5 (4), 857-865 (1989)
MEDLINE 90077427
PUBMED 2591967
REFERENCE 4 (bases 1 to 1506)
AUTHORS Pereira F, Rosenmann E, Nylen E, Kaufman M, Pinsky L and Wrogemann K.
TITLE The 56 kDa androgen binding protein is an aldehyde dehydrogenase
JOURNAL Biochem. Biophys. Res. Commun. 175 (3), 831-838 (1991)
MEDLINE 91222190
PUBMED 1709013
REFERENCE 5 (bases 1 to 1506)
AUTHORS Zheng,C.F., Wang,T.T. and Weiner,H.
TITLE Cloning and expression of the full-length cDNAS encoding human liver class 1 and class 2 aldehyde dehydrogenase
JOURNAL Alcohol. Clin. Exp. Res. 17 (4), 828-831 (1993)
MEDLINE 94027752
REFERENCE 6 (bases 1 to 1506)
AUTHORS Kathmann,E.C. and Lipsky,J.J.
TITLE Cloning of a cDNA encoding a constitutively expressed rat liver

FIG. 11 (2/3)

```
            cytosolic aldehyde dehydrogenase
  JOURNAL   Biochem. Biophys. Res. Commun. 236 (2), 527-531 (1997)
  MEDLINE   97382470
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from AF003341.1.
FEATURES             Location/Qualifiers
     source          1..1506
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /map="9q21"
                     /tissue_type="liver"
     gene            1..1506
                     /gene="ALDH1"
                     /note="PUMB1"
                     /db_xref="LocusID:216"
                     /db_xref="MIM:100640"
     CDS             1..1506
                     /gene="ALDH1"
                     /EC_number="1.2.1.3"
                     /note="cytosolic protein; class 1"
                     /codon_start=1
                     /db_xref="LocusID:216"
                     /db_xref="MIM:100640"
                     /product="aldehyde dehydrogenase 1, soluble"
                     /protein_id="NP_000680.1"
                     /db_xref="GI:4502031"
                     /translation="MSSSGTPDLPVLLTDLKIQYTKIFINNEWHDSVSGKKFPVFNPA
                     TEEELCQVEEGDKEDVDKAVKAARQAFQIGSPWRTMDASERGRLLYKLADLIERDRLL
                     LATMESMNGGKLYSNAYLSDLAGCIKTLRYCAGWADKIQGRTIPIDGNFFTYTRHEPI
                     GVCGQIIPWNFPLVMLIWKIGPALSCGNTVVVKPAEQTPLTALHVASLIKEAGFPPGV
                     VNIVPGYGPTAGAAISSHMDIDKVAFTGSTEVGKLIKEAAGKSNLKRVTLELGGKSPC
                     IVLADADLDNAVEFAHHGVFYHQGQCCIAASRIFVEESIYDEFVRRSVERAKKYILGN
                     PLTPGVTQGPQIDKEQYDKILDLIESGKKEGAKLECGGGPWGNKGYFVQPTVFSNVTD
                     EMRIAKEEIFGPVQQIMKFKSLDDVIKRANNTFYGLSAGVFTKDIDKAITISSALQAG
                     TVWVNCYGVVSAQCPFGGFKMSGNGRELGEYGFHEYTEVKTVTVKISQKNS"
     misc_feature    82..1488
                     /note="aldedh; Region: Aldehyde dehydrogenase family"
     variation       362
                     /allele="A"
                     /allele="G"
                     /db_xref="dbSNP:1049981"
     variation       1337
                     /allele="A"
                     /allele="C"
                     /db_xref="dbSNP:1803054"
     variation       1397
                     /allele="A"
                     /allele="T"
                     /db_xref="dbSNP:1063447"
BASE COUNT      441 a    293 c    391 g    381 t
ORIGIN
        1 atgtcatcct caggcacgcc agacttacct gtcctactca ccgatttgaa gattcaatat
       61 actaagatct tcataaacaa tgaatggcat gattcagtga gtggcaagaa atttcctgtc
      121 tttaatcctg caactgagga ggagctctgc caggtagaag aaggagataa ggaggatgtt
      181 gacaaggcag tgaaggccgc aagacaggct tttcagattg gatctccgtg gcgtactatg
      241 gatgcttccg agagggggcg actattatac aagttggctg atttaatcga aagagatcgt
      301 ctgctgctgg cgacaatgga gtcaatgaat ggtggaaaac tctattccaa tgcatatctg
      361 agtgatttag caggctgcat caaaacattg cgctactgtg caggttgggc tgacaagatc
      421 cagggccgta caataccaat tgatggaaat ttttttacat atacaagaca tgaacctatt
```

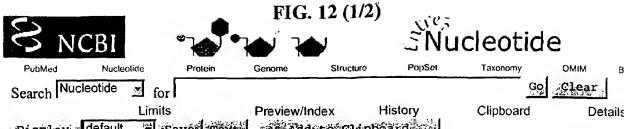
FIG. 11 (3/3)

```
 481 ggtgtatgtg gccaaatcat tccttggaat ttcccgttgg ttatgctcat ttggaagata
 541 gggcctgcac tgagctgtgg aaacacagtg gttgtcaaac cagcagagca aactcctctc
 601 actgctctcc acgtggcatc tttaataaaa gaggcagggt ttcctcctgg agtagtgaat
 661 attgttcctg gttatgggcc tacagcaggg gcagccattt cttctcacat ggatatagac
 721 aaagtagcct tcacaggatc aacagaggtt ggcaagttga tcaaagaagc tgccgggaaa
 781 agcaatctga agagggtgac cctggagctt ggaggaaaga gcccttgcat tgtgttagct
 841 gatgccgact tggacaatgc tgttgaattt gcacaccatg gggtattcta ccaccagggc
 901 cagtgttgta tagccgcatc caggattttt gtggaagaat caatttatga tgagtttgtt
 961 cgaaggagtg ttgagcgggc taagaagtat atccttggaa atcctctgac cccaggagtc
1021 actcaaggcc ctcagattga caaggaacaa tatgataaaa tacttgacct cattgagagt
1081 gggaagaaag aaggggccaa actggaatgt ggaggaggcc cgtgggggaa taaaggctac
1141 tttgtccagc ccacagtgtt ctctaatgtt acagatgaga tgcgcattgc caaagaggag
1201 atttttggac cagtgcagca aatcatgaag tttaaatctt tagatgacgt gatcaaaaga
1261 gcaaacaata ctttctatgg cttatcagca ggagtgttta ccaaagacat tgataaagcc
1321 ataacaatct cctctgctct gcaggcagga acagtgtggg tgaattgcta tggcgtggta
1381 agtgcccagt gccccttttgg cggattcaag atgtctggaa atggaagaga actgggagag
1441 tacggtttcc atgaatatac agaggtcaaa acagtcacag tgaaaatctc tcagaagaac
1501 tcataa
```

//

Revised: October 24, 2001.

FIG. 12 (1/2)



PYRUVATE KINASE

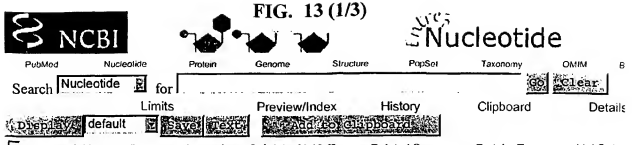
```
LOCUS       XM_037768               2282 bp    mRNA    linear   PRI 07-FEB-2002
DEFINITION  Homo sapiens similar to pyruvate kinase, muscle (H. sapiens)
            (LOC145710), mRNA.
ACCESSION   XM_037768
VERSION     XM_037768.1  GI:14750404
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2282)
  AUTHORS   NCBI Annotation Project.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2002) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
COMMENT     GENOME ANNOTATION REFSEQ:  This model reference sequence was
            predicted from NCBI contig NT_010235 by automated computational
            analysis using gene prediction method: BLAST. ~Also see:~
            Documentation of NCBI's Annotation Process~     Evidence Viewer ~
            alignments supporting this model.
FEATURES             Location/Qualifiers
     source          1..2282
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
     gene            1..2282
                     /gene="LOC145710"
                     /note="Located on Accession NT_010235"
                     /db_xref="InterimID:145710"
     CDS             109..1704
                     /gene="LOC145710"
                     /note="Located on Accession NT_010235"
                     /codon_start=1
                     /product="similar to pyruvate kinase, muscle (H. sapiens)"
                     /protein_id="XP_037768.1"
                     /db_xref="GI:14750405"
                     /translation="MSKPHSEAGTAFIQTQQLHAAMADTFLEHMCRLDIDSPPITARN
                     TGIICTIGPASRSVETLKEMIKSGMNVARLNFSHGTHEYHAETIKNVRTATESFASDP
                     ILYRPVAVALDTKGPEIRTGLIKGSGTAEVELKKGATLKITLDNAYMEKCDENILWLD
                     YKNICKVVEVGSKIYVDDGLISLQVKQKGADFLVTEVENGGSLGSKKGVNLPGAAVDL
                     PAVSEKDIQDLKFGVEQDVDMVFASFIRKASDVHEVRKVLGEKGKNIKIISKIENHEG
                     VRRFDEILEASDGIMVARGDLGIEIPAEKVFLAQKMMIGRCNRAGKPVICATQMLESM
                     IKKPRPTRAEGSDVANAVLDGADCIMLSGETAKGDYPLEAVRMQHLIAREAEAAIYHL
                     QLFEELRRLAPITSDPTEATAVGAVEASFKCCSGAIIVLTKSGRSAHQVARYRPRAPI
                     IAVTRNPQTARQAHLYRGIFPVLCKDPVQEAWAEDVDLRVNFAMNVGKARGFFKKGDV
                     VIVLTGWRPGSGFTNTMRVVPVP"
     misc_feature    223..1293
```

FIG. 12 (2/2)

```
                     /note="PK; Region: Pyruvate kinase, barrel domain"
     variation       546
                     /allele="C"
                     /allele="T"
                     /db_xref="dbSNP:10514"
     misc_feature    1333..1695
                     /note="PK_C; Region: Pyruvate kinase, alpha/beta domain"
     variation       2168
                     /allele="C"
                     /allele="T"
                     /db_xref="dbSNP:1062430"
BASE COUNT      499 a    646 c    654 g    483 t
ORIGIN
        1 ggctgaggca gtggctcctt gcacagcagc tgcacgcgcc gtggctccgg atctcttcgt
       61 ctttgcagcg tagcccgagt cggtcagcgc cggaggacct cagcagccat gtcgaagccc
      121 catagtgaag ccgggactgc cttcattcag acccagcagc tgcacgcagc catggctgac
      181 acattcctgg agcacatgtg ccgcctggac attgattcac caccatcac agcccggaac
      241 actggcatca tctgtaccat tggcccagct tcccgatcag tggagacgtt gaaggagatg
      301 attaagtctg gaatgaatgt ggctcgtctg aacttctctc atggaactca tgagtaccat
      361 gcggagacca tcaagaatgt gcgcacagcc acggaaagct ttgcttctga ccccatcctc
      421 taccggcccg ttgctgtggc tctagacact aaaggacctg agatccgaac tgggctcatc
      481 aagggcagcg gcactgcaga ggtggagctg aagaagggag ccactctcaa aatcacgctg
      541 gataacgcct acatggaaaa gtgtgacgag aacatcctgt ggctggacta caagaacatc
      601 tgcaaggtgg tggaagtggg cagcaagatc tacgtggatg atgggcttat ttctctccag
      661 gtgaagcaga aaggtgccga cttcctggtg acggaggtgg aaaatggtgg ctccttgggc
      721 agcaagaagg gtgtgaacct tcctggggct gctgtggact tgcctgctgt gtcggagaag
      781 gacatccagg atctgaagtt tggggtcgag caggatgttg atatggtgtt tgcgtcattc
      841 atccgcaagg catctgatgt ccatgaagtt aggaaggtcc tgggagagaa gggaaagaac
      901 atcaagatta tcagcaaaat cgagaatcat gagggggttc ggaggtttga tgaaatcctg
      961 gaggccagtg atgggatcat ggtggctcgt ggtgatctag gcattgagat tcctgcagag
     1021 aaggtcttcc ttgctcagaa gatgatgatt ggacggtgca accgagctgg gaagcctgtc
     1081 atctgtgcta ctcagatgct ggagagcatg atcaagaagc cccgccccac tcgggctgaa
     1141 ggcagtgatg tggccaatgc agtcctggat ggagccgact gcatcatgct gtctggagaa
     1201 acagccaaag gggactatcc tctggaggct gtgcgcatgc agcacctgat tgcccgtgag
     1261 gcagaggctg ccatctacca cttgcaatta tttgaggaac tccgccgcct ggcgcccatt
     1321 accagcgacc ccacagaagc caccgccgtg ggtgccgtgg aggcctcctt caagtgctgc
     1381 agtgggggcca taatcgtcct caccaagtct ggcaggtctg ctcaccaggt ggccagatac
     1441 cgcccacgtg cccccatcat tgctgtgacc cggaatcccc agacagctcg tcaggcccac
     1501 ctgtaccgtg gcatcttccc tgtgctgtgc aaggacccag tccaggaggc ctgggctgag
     1561 gacgtggacc tccgggtgaa ctttgccatg aatgttggca aggcccgagg cttcttcaag
     1621 aagggagatg tggtcattgt gctgaccgga tggcgccctg gctccggctt caccaacacc
     1681 atgcgtgttg ttcctgtgcc gtgatggacc ccagagcccc tcctccagcc cctgtcccac
     1741 cccccttcccc cagcccatcc attaggccag caacgcttgt agaactcact ctgggctgta
     1801 acgtggcact ggtaggttgg gacaccaggg aagaagatca acgcctcact gaaacatggc
     1861 tgtgtttgca gcctgctcta gtgggacagc ccagagcctg gctgcccatc atgtggcccc
     1921 acccaatcaa gggaagaagg aggaatgctg gactggaggc ccctggagcc agatggcaag
     1981 agggtgacag cttcctttcc tgtgtgtact ctgtccagtt cctttagaaa aaatggatgc
     2041 ccagaggact cccaaccctg gcttggggtc aagaaacagc cagcaagagt taggggcctt
     2101 agggcactgg gctgttgttc cattgaagcc gactctggcc ctggccctta cttgcttctc
     2161 tagctctcta ggcctctcca gtttgcacct gtccccaccc tccactcagc tgtcctgcag
     2221 caaacactcc accctccacc ttccattttc ccccactact gcagcacctc caggcctgtt
     2281 gc
//
```

Revised: October 24, 2001.

FIG. 13 (1/3)



LOCUS XM_049337 2631 bp mRNA linear PRI 07-FEB-2002
DEFINITION Homo sapiens glucose-6-phosphate dehydrogenase (G6PD), mRNA.
ACCESSION XM_049337
VERSION XM_049337.1 GI:14768486
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2631)
AUTHORS NCBI Annotation Project.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-2002) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA
COMMENT GENOME ANNOTATION REFSEQ: This model reference sequence was predicted from NCBI contig NT_025965 by automated computational analysis using gene prediction method: BLAST. ~Also see:~ Documentation of NCBI's Annotation Process~ Evidence Viewer - alignments supporting this model.

G6PD

FEATURES Location/Qualifiers
source 1..2631
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
gene 1..2631
/gene="G6PD"
/note="G6PD1; Located on Accession NT_025965"
/db_xref="LocusID:2539"
/db_xref="MIM:305900"
CDS 475..2022
/gene="G6PD"
/note="Located on Accession NT_025965"
/codon_start=1
/product="glucose-6-phosphate dehydrogenase"
/protein_id="XP_049337.1"
/db_xref="GI:14768487"
/translation="MAEQVALSRTQVCGILREELFQGDAFHQSDTHIFIIMGASGDLA
KKKIYPTIWWLFRDGLLPENTFIMGYARSRLTVADIRKQSEPFFKATPEEKLKLEDFF
ARNSYVAGQYDDAASYQRLNSHMDALHLGSQANRLFYLALPPTVYEAVTKNIHESCMS
QIGWNRIIVEKPFGRDLQSSDRLSNHISSLFREDQIYRIDHYLGKEMVQNLMVLRFAN
RIFGPIWNRDNIACVILTFKEPFGTEGRGGYFDEFGIIRDVMQNHLLQMLCLVAMEKP
ASTNSDDVRDEKVKVLKCISEVQANNVVLGQYVGNPDGEGEATKGYLDDPTVPRGSTT
ATFAAVVLYVENERWDGVPFILRCGKALNERKAEVRLQFHDVAGDIFHQQCKRNELVI
RVQPNEAVYTKMMTKKPGMFFNPEESELDLTYGNRYKNVKLPDAYERLILDVFCGSQM
HFVRSDELREAWRIFTPLLHQIELEKPKPIPYIYGSRGPTEADELMKRVGFQYEGTYK
WVNPHKL"
variation 507

FIG. 13 (2/3)

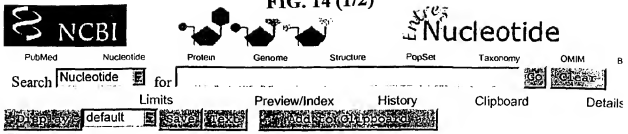
```
                     /allele="C"
                     /allele="G"
                     /db_xref="dbSNP:1050827"
     misc_feature    553..1104
                     /note="G6PD; Region: Glucose-6-phosphate dehydrogenase,
                     NAD binding domain"
     variation       676
                     /allele="A"
                     /allele="G"
                     /db_xref="dbSNP:1050828"
     variation       850
                     /allele="A"
                     /allele="G"
                     /db_xref="dbSNP:1050829"
     misc_feature    1108..1992
                     /note="G6PD_C; Region: Glucose-6-phosphate dehydrogenase,
                     C-terminal domain"
     variation       2379
                     /allele="A"
                     /allele="G"
                     /db_xref="dbSNP:1050757"
     variation       2392
                     /allele="A"
                     /allele="G"
                     /db_xref="dbSNP:1063529"
     variation       2490
                     /allele="A"
                     /allele="G"
                     /db_xref="dbSNP:1050830"
     variation       2553
                     /allele="C"
                     /allele="T"
                     /db_xref="dbSNP:1050773"
     variation       2555
                     /allele="C"
                     /allele="T"
                     /db_xref="dbSNP:1050774"
     variation       2593
                     /allele="C"
                     /allele="T"
                     /db_xref="dbSNP:1050831"
BASE COUNT       527 a    884 c    797 g    423 t
ORIGIN
        1 agggacagcc cagaggaggc gtggccacgc tgccggcgga agtggagccc tccgcgagcg
       61 cgcgaggccg ccggggcagg cggggaaacc ggacagtagg ggcggggccg ggccggcgat
      121 ggggatgcgg gagcactacg cggagctgca cccgtgcccg ccggaattgg ggatgcagag
      181 cagcggcagc gggtatggca ggcagccggc gggccggcct ccagcgcagg tgcccgagag
      241 gcaggggctg gcctgggatg cgcgcgcacc tgccctcgcc ccgccccgcc cgcacgaggg
      301 gtggtggccg aggccccgcc ccgcacgcct cgcctgaggc gggtccgctc agcccaggcg
      361 cccgcccccg ccccgccgca ttaaatgggc cggcggggct cagccccgg aaacggtcgt
      421 acacttcggg gctgcgagcg cggagggcga cgacgacgaa gcgcagacag cgtcatggca
      481 gagcaggtgg ccctgagccg gacccaggtg tgcgggatcc tgcgggaaga gcttttccag
      541 ggcgatgcct tccatcagtc ggatacacac atattcatca tcatgggtgc atcggggtgac
      601 ctggccaaga agaagatcta ccccaccatc tggtggctgt tccgggatgg ccttctgccc
      661 gaaaacacct tcatcatggg ctatgcccgt tcccgcctca cagtggctga catccgcaaa
      721 cagagtgagc ccttcttcaa ggccacccca gaggagaagc tcaagctgga ggacttcttt
      781 gcccgcaact cctatgtggc tggccagtac gatgatgcag cctcctacca gcgcctcaac
      841 agccacatgg atgccctcca cctggggtca caggccaacc gcctcttcta cctggccttg
      901 cccccgaccg tctacgaggc cgtcaccaag aacattcacg agtcctgcat gagccagata
      961 ggctggaacc gcatcatcgt ggagaagccc ttcgggaggg acctgcagag ctctgaccgg
```

FIG. 13 (3/3)

```
1021 ctgtccaacc acatctcctc cctgttccgt gaggaccaga tctaccgcat cgaccactac
1081 ctgggcaagg agatggtgca gaacctcatg gtgctgagat ttgccaacag gatcttcggc
1141 cccatctgga accgggacaa catcgcctgc gttatcctca ccttcaagga gcccttggc
1201 actgagggtc gcgggggcta tttcgatgaa tttgggatca tccgggacgt gatgcagaac
1261 cacctactgc agatgctgtg tctggtggcc atggagaagc ccgcctccac caactcagat
1321 gacgtccgtg atgagaaggt caaggtgttg aaatgcatct cagaggtgca ggccaacaat
1381 gtggtcctgg gccagtacgt ggggaacccc gatggagagg gcgaggccac caaagggtac
1441 ctggacgacc ccacggtgcc ccgcgggtcc accaccgcca cttttgcagc cgtcgtcctc
1501 tatgtggaga atgagaggtg ggatggggtg cccttcatcc tgcgctgcgg caaggccctg
1561 aacgagcgca aggccgaggt gaggctgcag ttccatgatg tggccggcga catcttccac
1621 cagcagtgca agcgcaacga gctggtgatc cgcgtgcagc ccaacgaggc cgtgtacacc
1681 aagatgatga ccaagaagcc gggcatgttc ttcaaccccg aggagtcgga gctggacctg
1741 acctacggca acagatacaa gaacgtgaag ctccctgacg cctacgagcg cctcatcctg
1801 gacgtcttct gcggggagcca gatgcacttc gtgcgcagcg acgagctccg tgaggcctgg
1861 cgtatttttca ccccactgct gcaccagatt gagctggaga agcccaagcc catccctat
1921 atttatggca gccgaggccc cacggaggca gacgagctga tgaagagagt gggtttccag
1981 tatgagggca cctacaagtg ggtgaacccc cacaagctct gagccctggg cacccacctc
2041 caccccgcc acggccaccc tccttcccgc cgcccgaccc cgagtcggga ggactccggg
2101 accattgacc tcagctgcac attcctggcc ccgggctctg gccaccctgg cccgccctc
2161 gctgctgcta ctaccgagc ccagctacat tcctcagctg ccaagcactc gagaccatcc
2221 tggccctcc agaccctgcc tgagcccagg agctgagtca cctcctcdac tcactccagc
2281 ccaacagaag gaaggaggag ggcgcccatt cgtctgtccc agagcttatt ggccactggg
2341 tctcactcct gagtggggcc agggtgggag ggagggacaa gggggaggaa agggggcgagc
2401 acccacgtga gagaatctgc ctgtggcctt gcccgccagc ctcagtgcca cttgacattc
2461 cttgtcacca gcaacatctc gagccccctg gatgtcccct gtcccaccaa ctctgcactc
2521 catggccacc ccgtgccacc cgtaggcagc ctctctgcta taagaaaagc agacgcagca
2581 gctgggaccc ctcccaacct caatgccctg ccattaaatc cgcaaacagc c
//
```

Revised: October 24, 2001.

FIG. 14 (1/2)



1: XM_049047[gi:14759750]

```
LOCUS       XM_049047                1564 bp    mRNA    linear   PRI 16-JUL-2001
DEFINITION  Homo sapiens proliferation-associated 2G4, 38kD (PA2G4), mRNA.
ACCESSION   XM_049047
VERSION     XM_049047.1  GI:14759750
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1564)
  AUTHORS   NCBI Annotation Project.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-2001) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
FEATURES             Location/Qualifiers
     source          1..1564
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
     gene            1..1564
                     /gene="PA2G4"
                     /db_xref="LocusID:5036"
                     /db_xref="MIM:602145"
     CDS             120..1304
                     /gene="PA2G4"
                     /codon_start=1
                     /product="proliferation-associated 2G4, 38kD"
                     /protein_id="XP_049047.1"
                     /db_xref="GI:14759751"
                     /translation="MSGEDEQQEQTIAEDLVVTKYKMGGDIANRVLRSLVEASSSGVS
                     VLSLCEKGDAMIMEETGKIFKKEKEMKKGIAFPTSISVNNCVCHFSPLKSDQDYILKE
                     GDLVKIDLGVHVDGFIANVAHTFVVDVAQGTQVTGRKADVIKAAHLCAEAALRLVKPG
                     NQNTQVTEAWNKVAHSFNCTPIEGMLSHQLKQHVIDGEKTIIQNPTDQQKKDHEKAEF
                     EVHEVYAVDVLVSSGEGKAKDAGQRTTIYKRDPSKQYGLKMKTSRAFFSEVERRFDAM
                     PFTLRAFEDEKKARMGVVECAKHELLQPFNVLYEKEGEFVAQFKFTVLLMPNGPMRIT
                     SGPFEPDLYKSEMEVQDAELKALLQSSASRKTQKKKKKKASKTAENATSGETLEENEA
                     GD"
BASE COUNT      455 a    365 c    413 g    331 t
ORIGIN
        1 ctttcgctcg ccctctcctc gaggatcgag gggactctga ccacagcctg tggctgggaa
       61 gggagacaga ggcggcggcg gctcagggga aacgaggctg cagtggtggt agtaggaaga
      121 tgtcgggcga ggacgagcaa caggagcaaa ctatcgctga ggacctggtc gtgaccaagt
      181 ataagatggg gggcgacatc gccaacaggg tacttcggtc cttggtggaa gcatctagct
      241 caggtgtgtc ggtactgagc ctgtgtgaga aaggtgatgc catgattatg gaagaaacag
      301 ggaaaatctt caagaaagaa aaggaaatga agaaaggtat tgcttttccc accagcattt
      361 cggtaaataa ctgtgtatgt cacttctccc ctttgaagag cgaccaggat tatattctca
      421 aggaaggtga cttggtaaaa attgaccttg gggtccatgt ggatggcttc atcgctaatg
```

HCDR-3

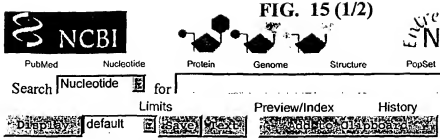
FIG. 14 (2/2)

```
 481 tagctcacac ttttgtggtt gatgtagctc aggggaccca agtaacaggg aggaaagcag
 541 atgttattaa ggcagctcac ctttgtgctg aagctgccct acgcctggtc aaacctggaa
 601 atcagaacac acaagtgaca gaagcctgga acaaagttgc ccactcattt aactgcacgc
 661 caatagaagg tatgctgtca caccagttga agcagcatgt catcgatgga gaaaaaacca
 721 ttatccagaa tcccacagac cagcagaaga aggaccatga aaaagctgaa tttgaggtac
 781 atgaagtata tgctgtggat gttctcgtca gctcaggaga gggcaaggcc aaggatgcag
 841 gacagagaac cactatttac aaacgagacc cctctaaaca gtatggactg aaaatgaaaa
 901 cttcacgtgc cttcttcagt gaggtggaaa ggcgttttga tgccatgccg tttacttttaa
 961 gagcatttga agatgagaag aaggctcgga tgggtgtggt ggagtgcgcc aaacatgaac
1021 tgctgcaacc atttaatgtt ctctatgaga aggagggtga atttgttgcc cagtttaaat
1081 ttacagttct gctcatgccc aatggcccca tgcggataac cagtggtccc ttcgagcctg
1141 acctctacaa gtctgagatg gaggtccagg atgcagagct aaaggccctc ctccagagtt
1201 ctgcaagtcg aaaaacccag aaaaagaaaa aaaagaaggc ctccaagact gcagagaatg
1261 ccaccagtgg ggaaacatta gaagaaaatg aagctgggga ctgaggtggg tcccatctcc
1321 ccagcttgct gctcctgcct catccccttc ccaccaaacc ccagactctg tgaagtgcag
1381 ttcttctcca cctaggaccg ccagcagagc ggggggatct ccctgccccc accccagttc
1441 cccaacccac tcccttccaa caacaaccag ctccaactga ctctggtctt gggaggtgag
1501 gcttcccaac cacggaagac tactttaaat gaaaaaaaga aattgaataa taaaatcagg
1561 agtc
```

//

Revised: October 24, 2001.

FIG. 15 (1/2)



1: XM_052326[gi:14748477]

DDX21

```
LOCUS       XM_052326                3273 bp    mRNA    linear   PRI 16-JUL-2001
DEFINITION  Homo sapiens DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 21
            (DDX21), mRNA.
ACCESSION   XM_052326
VERSION     XM_052326.1  GI:14748477
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3273)
  AUTHORS   NCBI Annotation Project.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-2001) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
FEATURES             Location/Qualifiers
     source          1..3273
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
     gene            1..3273
                     /gene="DDX21"
                     /note="GURDB; RH-II/GU"
                     /db_xref="LocusID:9188"
     CDS             35..1711
                     /gene="DDX21"
                     /codon_start=1
                     /product="DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 21"
                     /protein_id="XP_052326.1"
                     /db_xref="GI:14748478"
                     /translation="MPGKLRSDAGLESDTAMKKGETLRKQTEEKEKKEKPKSDKTEEI
                     AEEEETVPPKAKQVKKKAEPSEVDMNSPKSKKAKKKEEPSQNDISPKTKSLRKKKEPI
                     EKKVVSSKTKKVTKNEEPSEEEIDAPKPKKMKKEKEMNGETREKSPKLKNGFPHPEPD
                     CNPSEAASEESNSEIEQEIPVEQKEGAFSNFPISEETIKLLKGRGVTFLFPIQAKTFH
                     HVYSGKDLIAQARTGTGKTFSFAIPLIEKLHGELQDRKRGRAPQVLVLAPTRELANQV
                     SKDFSDITKKLSVACFYGGTPYGGQFERMRNGIDILVGTPGRIKDHIQNGKLDLTKLK
                     HVVLDEVDQMLDMGFADQVEEILSVAYKKDSEDNPQTLLFSATCPHWVFNVAKKYMKS
                     TYEQVDLIGKKTQKTAITVEHLAIKCHWTQRAAVIGDVIRVYSGHQGRTIIFCETKKE
                     AQELSQNSAIKQDAQSLHGDIPQKQREITLKGFRNGSFGVLVATNVAARGLDIPEVDL
                     VIQSSPPKGCRVLHSSIRADRQSWKDGGVHLLLSAQGRISVSTSGAKSGN"
BASE COUNT     1068 a    603 c    773 g    829 t
ORIGIN
        1 gaagaccggt cggcctgggc aacctgcgct gaagatgccg ggaaaactcc gtagtgacgc
       61 tggtttggaa tcagacaccg caatgaaaaa aggggagaca ctgcgaaagc aaaccgagga
      121 gaaagagaaa aaagagaagc caaaatctga taagactgaa gagatagcag aagaggaaga
      181 aactgttttc cccaaagcta aacaagttaa aaagaaagca gagccttctg aagttgacat
      241 gaattctcct aaatccaaaa aggcaaaaaa gaaagaggag ccatctcaaa atgacatttc
```

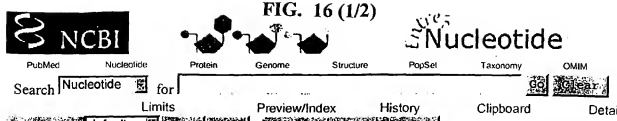
FIG. 15 (2/2)

```
 301 tcctaaaacc aaaagtttga gaaagaaaaa ggagccdatt gaaaagaaag tggtttcttc
 361 taaaaccaaa aaagtgacaa aaaatgagga gccttctgag gaagaaatag atgctcctaa
 421 gcccaagaag atgaagaaag aaaaggaaat gaatggagaa actagagaga aaagccccaa
 481 actgagaaat ggatttcctc atcctgaacc ggactgtaac cccagtgaag ctgccagtga
 541 agaaagtaac agtgagatag agcaggaaat acctgtggaa caaaaagaag gcgctttctc
 601 taatttttcc atatctgaag aaactattaa acttctcaaa ggccgaggag tgaccttcct
 661 atttcctata caagcaaaga cattccatca tgtttacagc gggaaggact taattgcaca
 721 ggcacggaca ggaactggga agacattctc ctttgccatc cctttgattg agaaacttca
 781 tggggaactg caagacagga agagaggccg tgcccctcag gtactggttc ttgcacctac
 841 aagagagttg gcaaatcaag taagcaaaga cttcagtgac atcacaaaaa agctgtcagt
 901 ggcttgtttt tatggtggaa ctccctatgg aggtcaattt gaacgcatga ggaatgggat
 961 tgatatcctg gttggaacac caggtcgtat caaagaccac atacgaatg gcaaactaga
1021 tctcaccaaa cttaagcatg ttgtcctgga tgaagtggac cagatgttgg atatgggatt
1081 tgctgatcaa gtggaagaga ttttaagtgt ggcatacaag aaagattctg aagacaatcc
1141 ccaaacattg ctttttctg caacttgccc tcattgggta tttaatgttg ccaagaaata
1201 catgaaatct acatatgaac aggtggacct gattggtaaa aagactcaga aaacggcaat
1261 aactgtggag catctggcta ttaagtgcca ctggactcag agggcagcag ttattgggga
1321 tgtcatccga gtatatagtg gtcatcaagg acgcactatc atcttttgtg aaaccaagaa
1381 agaagcccag gagctgtccc agaattcagc tataaagcag gatgctcagt ccttgcatgg
1441 agacatccca cagaagcaaa gggaaatcac cctgaaaggt ttagaaatg gtagttttgg
1501 agttttggtg gcaaccaatg ttgctgcacg tgggttagac atccctgagg ttgatttggt
1561 tatacaaagc tctccaccaa agggatgtag agtcctacat tcatcgatcc gggcggacag
1621 gcagagctgg aaggacgggg gtgtgcatct gcttttatca gcacaaggaa gaatatcagt
1681 tagtacaagt ggagcaaaaa gcgggaatta agttcaaacg aataggtgtt ccttctgcaa
1741 cagaaataat aaaagcttcc agcaaagatg ccatcaggct tttggattcc gtgcctccca
1801 ctgccattag tcacttcaaa caatcagctg agaagctgat agaagcagag ggagctgtgg
1861 aagctctggc agcagcactg gcccatattt caggtgccac gtccgtagac cagcgctcct
1921 tgatcaactc aaatgtgggt tttgtgacca tgatcttgca gtgctcaatt gaaatgccaa
1981 atattagtta tgcttggaaa gaacttaaag agcagctggg cgaggagatt gattccaaag
2041 tgaaggaagt ggttttctc aaaggaaagc tgggtgtttg ctttgatgta cctaccgcat
2101 cagtaacaga aatacaggag aaatggcatg attcacgacg ctggcagctc tctgtggcca
2161 cagagcaacc agaactggaa ggaccacggg aaggatatgg aggcttcagg ggacagcggg
2221 aaggcagtcg aggcttcagg ggacagcggg acggaaacag aagattcaga ggacagcggg
2281 aaggcagtag aggccgaga ggacagcgat caggaggtgg caacaaaagt aacagatccc
2341 aaaacaaagg ccagaagcgg agtttcagta aagcatttgg tcaataatta gaaatagaag
2401 attatatag caaaaagaga atgatgtttg gcaatataga acgaacatt attttcatg
2461 caaagttaaa agcacattgt gcctcctttt gaccacttgc caagtccctg tctctttcag
2521 acacagacaa gcttcattta aattatttca tctgatcatt atcatttata acttttattg
2581 tacttcttca tcagtttttc cttttgaaag gtgtatgaat tcattacttt tttattctaa
2641 tgtattatct gtagattaga agataaaatc aagcattgtc tgcctatac tttgtgagtt
2701 cacctgtctt tatactcaaa agtgtccctt aatagtgtcc ttccctgaaa taaatacctа
2761 agggagtgta acagtctctg gaggaccact ttgagccttt ggaagttaag gtttcctcag
2821 ccacctgccg aacagtttct catgtggtcc tattatttgt ctactgagac ttaatactga
2881 gcaatgtttt gaaacaagat ttcaaactaa tctgggttgt aatacagttt atacgagtgt
2941 atgctctaga cttggaagat gtagtatgtt tgatgtggat tacctatact tatgttcgtt
3001 ttgatacatt tttagcttct cattataagg tgattcatgc ttctgtgaat tcttcataga
3061 tagtatatat aaaagtacat tttaatagaa agccagggtt ttaaggaatt tcacatgtat
3121 aaggtggctc catagcttta tttgtaagta ggctggataa atggtgctta aatggtaatg
3181 tactccactt cttcctattg gaagattaac attatttacc aagaaggact taagggagta
3241 gggggcgcag attagcattg ctcaagagta tgt
```

//

Revised: October 24, 2001.

FIG. 16 (1/2)



1: XM_030607[gi:14786409]

ARK2

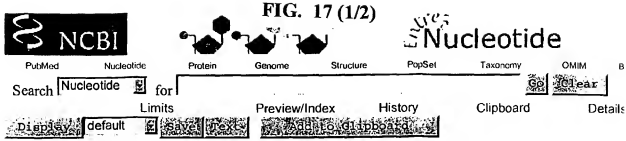
```
LOCUS       XM_030607               2005 bp    mRNA    linear   PRI 16-OCT-2001
DEFINITION  Homo sapiens serine/threonine kinase 15 (STK15), mRNA.
ACCESSION   XM_030607
VERSION     XM_030607.1  GI:14786409
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2005)
  AUTHORS   NCBI Annotation Project.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2001) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
FEATURES             Location/Qualifiers
     source          1..2005
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
     gene            1..2005
                     /gene="STK15"
                     /note="BTAK; Located on Accession NT_011362"
                     /db_xref="LocusID:8465"
                     /db_xref="MIM:603072"
     CDS             26..1237
                     /gene="STK15"
                     /note="Located on Accession NT_011362"
                     /codon_start=1
                     /product="hypothetical protein XP_030607"
                     /protein_id="XP_030607.1"
                     /db_xref="GI:14786410"
                     /translation="MDRSKENCISGPVKATAPVGGPKRVLVTQQFPCQNPLPVNSGQA
                     QRVLCPSNSSQRIPLQAQKLVSSHKPVQNQKQKQLQATSVPHPVSRPLNNTQKSKQPL
                     PSAPENNPEEELASKQKNEESKKRQWALEDFEIGRPLGKGKFGNVYLAREKQSKFILA
                     LKVLFKAQLEKAGVEHQLRREVEIQSHLRHPNILRLYGYFHDATRVYLILEYAPLGTV
                     YRELQKLSKFDEQRTATYITELANALSYCHSKRVIHRDIKPENLLLGSAGELKIADFG
                     WSVHAPSSRRTTLCGTLDYLPPEMIEGRMHDEKVDLWSLGVLCYEFLVGKPPFEANTY
                     QETYKRISRVEFTFPDFVTEGARDLISRLLKHNPSQRPMLREVLEHPWITANSSKPSN
                     CQNKESASKQS"
     misc_feature    422..1174
                     /note="S_TKc; Region: Serine/Threonine protein kinases,
                     catalytic domain"
     misc_feature    422..1174
                     /note="pkinase; Region: Protein kinase domain"
     misc_feature    425..1162
                     /note="TyrKc; Region: Tyrosine kinase, catalytic domain"
BASE COUNT      585 a    434 c    456 g    530 t
```

FIG. 16 (2/2)

```
ORIGIN
      1 cttgggtcct tgggtcgcag gcatcatgga ccgatctaaa gaaaactgca tttcaggacc
     61 tgttaaggct acagctccag ttggaggtcc aaaacgtgtt ctcgtgactc agcaatttcc
    121 ttgtcagaat ccattacctg taaatagtgg ccaggctcag cgggtcttgt gtccttcaaa
    181 ttcttcccag cgcattcctt tgcaagcaca aaagcttgtc tccagtcaca agccggttca
    241 gaatcagaag cagaagcaat tgcaggcaac cagtgtacct catcctgtct ccaggccact
    301 gaataacacc caaaagagca agcagcccct gccatcggca cctgaaaata atcctgagga
    361 ggaactggca tcaaaacaga aaaatgaaga atcaaaaaag aggcagtggg ctttggaaga
    421 cttgaaatt ggtcgccctc tgggtaaagg aaagtttggt aatgtttatt tggcaagaga
    481 aaagcaaagc aagtttattc tggctcttaa agtgttattt aaagctcagc tggagaaagc
    541 cggagtggag catcagctca gaagagaagt agaaatacag tcccaccttc ggcatcctaa
    601 tattcttaga ctgtatggtt atttccatga tgctaccaga gtctacctaa ttctggaata
    661 tgcaccactt ggaacagttt atagagaact tcagaaactt tcaaagtttg atgagcagag
    721 aactgctact tatataacag aattggcaaa tgccctgtct tactgtcatt cgaagagagt
    781 tattcataga gacattaagc cagagaactt acttcttgga tcagctggag agcttaaaat
    841 tgcagatttt gggtggtcag tacatgctcc atcttccagg aggaccactc tctgtggcac
    901 cctggactac ctgccccctg aaatgattga aggtcggatg catgatgaga aggtggatct
    961 ctggagcctt ggagttcttt gctatgaatt tttagttggg aagcctcctt ttgaggcaaa
   1021 cacatacaaa gagacctaca aaagaatatc acgggttgaa ttcacattcc ctgactttgt
   1081 aacagaggga gccagggacc tcatttcaag actgttgaag cataatccca gccagaggcc
   1141 aatgctcaga gaagtacttg aacacccctg gatcacagca aattcatcaa aaccatcaaa
   1201 ttgccaaaac aaagaatcag ctagcaaaca gtcttaggaa tcgtgcaggg ggagaaatcc
   1261 ttgagccagg gctgccatat aacctgacag gaacatgcta ctgaagttta ttttacactt
   1321 gactgctgcc ctcaatctag aacgctacac aagaaatatt tgttttactc agcaggtgtg
   1381 ccttaacctc cctattcaga aagctccaca tcaataaaca tgacactctg aagtgaaagt
   1441 agccacgaga attgtgctac ttatactggt tcataatctg gaggcaaggt tcgactgcag
   1501 ccgccccgtc agcctgtgct aggcatggtg tcttcacagg aggcaaatcc agagcctggc
   1561 tgtgggggaaa gtgaccactc tgccctgacc ccgatcagtt aaggagctgt gcaataacct
   1621 tcctagtacc tgagtgagtg tgtaacttat tgggtggcg aagcctggta aagctgttgg
   1681 aatgagtatg tgatttcttt taagtatgaa aataaagata tatgtacaga cttgtatttt
   1741 ttctctggtg gcattccttt aggaatgctg tgtgtctgtc cggcaccccg gtaggcctga
   1801 ttgggtttct agtcctcctt aaccacttat ctcccatatg agagtgtgaa aaataggaac
   1861 acgtgctcta cctccattta gggatttgct tgggatacag aagaggccat gtgtctcaga
   1921 gctgttaagg gcttattttt ttaaaacatt ggagtcatag catgtgtgta aactttaaat
   1981 atgcaaataa ataagtatct atgtc
//
```

Revised: October 24, 2001.

FIG. 17 (1/2)



LOCUS BC008442 1584 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, Similar to transmembrane 4 superfamily member 1, clone MGC:14656 IMAGE:4101110, mRNA, complete cds.
ACCESSION BC008442
VERSION BC008442.1 GI:14250074
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1584)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (25-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 21 Row: l Column: 7
This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

FEATURES Location/Qualifiers
source 1..1584
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:14656 IMAGE:4101110"
/tissue_type="Bone marrow, chronic myelogenous leukemia"
/clone_lib="NIH_MGC_54"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"
CDS 102..710

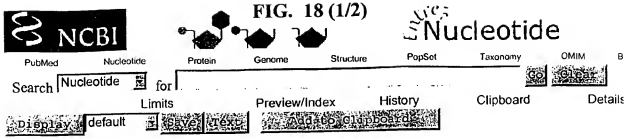
TM4 SF1

FIG. 17 (2/2)

```
                     /codon_start=1
                     /product="Similar to transmembrane 4 superfamily member 1"
                     /protein_id="AAH08442.1"
                     /db_xref="GI:14250075"
                     /translation="MCYGKCARCIGHSLVGLALLCIAANILLYFPNGETKYASENHLS
                     RFVWFFSGIVGGGLLMLLPAFVFIGLEQDDCCGCCGHENCGKRCAMLSSVLAALIGIA
                     GSGYCVIVAALGLAEGPLCLDSLGQWNYTFASTEGQYLLDTSTWSECTEPKHIVEWNV
                     SLFSILLALGGIEFILCLIQVINGVLGGICGFCCSHQQQYDC"
BASE COUNT       460 a    311 c    337 g    476 t
ORIGIN
        1 gtggtgtttg ctttctccac cagaagggca cactttcatc taatttgggg tatcactgag
       61 ctgaagacaa agagaagggg gagaaaacct agcagaccac catgtgctat gggaagtgtg
      121 cacgatgcat cggacattct ctggtggggc tcgccctcct gtgcatcgcg gctaatattt
      181 tgctttactt tcccaatggg gaaacaaagt atgcctccga aaaccacctc agccgcttcg
      241 tgtggttctt ttctggcatc gtaggaggtg gcctgctgat gctcctgcca gcatttgtct
      301 tcattgggct ggaacaggat gactgctgtg gctgctgtgg ccatgaaaac tgtggcaaac
      361 gatgtgcgat gctttcttct gtattggctg ctctcattgg aattgcagga tctggctact
      421 gtgtcattgt ggcagccctt ggcttagcag aaggaccact atgtcttgat tccctcggcc
      481 agtggaacta caccctttgcc agcaccgagg gccagtacct tctggatacc tccacatggt
      541 ccgagtgcac tgaacccaag cacattgtgg aatggaatgt atctctgttt tctatcctct
      601 tggctcttgg tggaattgaa ttcatcttgt gtcttattca agtaatattt ggagtgcttg
      661 gaggcatatg tggcttttgc tgctctcacc aacagcaata tgactgctaa aagaaccaac
      721 ccaggacaga gccacaatct tcctctattt cattgtaatt tatatatttc acttgtattc
      781 atttgtaaaa ctttgtatta gtgtaacata ctccccacag tctacttttа caaacgcctg
      841 taaagactgg catcttcaca ggatgtcagt gtttaaattt agtaaacttc ttttttgttt
      901 gtttatttgt ttttgttttt tttttaggaa tgaggaaaca aaccaccctc tgggggtagt
      961 ttacagactg agtgacagta ctcagtatat ctgagataaa ctctatаatg ttttgataa
     1021 aaataacatt ccaatcacta ttgtatatat gtgcatgtat tttttaaatt aaagatgtct
     1081 agttgctttt tataagacca agaaggagaa aatccgacaa cctggaaaga tttttgtttt
     1141 cactgcttgt atgatgtttc ccattcatac acctataaat ctctaacaag aggccctttg
     1201 aactgccttg tgttctgtga gaaacaaata ttтacttaga gtggaaggac tgattgagaa
     1261 tgttccaatc caaatgaatg catcacaact tacaatgctg ctcattgttg tgagtactat
     1321 gagattcaaa tttttctaac atatggaaag ccttttgtcc tccaaagatg agtactaggg
     1381 atcatgtgtt taaaaaaaag aaaggctacg atgactgggc aagaagaaag atgggaaact
     1441 gaataaagca gttgatcagc atcattggaa catggggacg agtgacggca ggaggaccac
     1501 gaggaaatac cctcaaaact aacttgttta caacaaaata aagtattcac tacgaaaaaa
     1561 aaaaaaaaaa aaaaaaaaa aaaa
//
```

Revised: October 24, 2001.

FIG. 18 (1/2)



1: XM_027538[gi:14768648]

ERCC1

```
LOCUS       XM_027538                1025 bp    mRNA    linear   PRI 16-JUL-2001
DEFINITION  Homo sapiens excision repair cross-complementing rodent repair
            deficiency, complementation group 1 (includes overlapping antisense
            sequence) (ERCC1), mRNA.
ACCESSION   XM_027538
VERSION     XM_027538.1  GI:14768648
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1025)
  AUTHORS   NCBI Annotation Project.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUL-2001) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
FEATURES             Location/Qualifiers
     source          1..1025
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
     gene            1..1025
                     /gene="ERCC1"
                     /note="UV20"
                     /db_xref="LocusID:2067"
                     /db_xref="MIM:126380"
     CDS             63..956
                     /gene="ERCC1"
                     /codon_start=1
                     /product="excision repair cross-complementing rodent
                     repair deficiency, complementation group 1 (includes
                     overlapping antisense sequence)"
                     /protein_id="XP_027538.1"
                     /db_xref="GI:14768649"
                     /translation="MDPGKDKEGVPQPSGPPARKKFVIPLDEDEVPPGVAKPLFRSTQ
                     SLPTVDTSAQAAPQTYAEYAISQPLEGAGATCPTGSEPLAGETPNQALKPGAKSNSII
                     VSPRQRGNPVLKFVRNVPWEFGDVIPDYVLGQSTCALFLSLRYHNLHPDYIHGRLQSL
                     GKNFALRVLLVQVDVKDPQQALKELAKMCILADCTLILAWSPEEAGRYLETYKAYEQK
                     PADLLMEKLEQDFVSRVTECLTTVKSVNKTDSQTLLTTFGSLEQLIAASREDLALCPG
                     LGPQKARRLFDVLHEPFLKVP"
BASE COUNT      234 a    326 c    289 g    176 t
ORIGIN
        1 ccaagaccag caggtgaggc ctcgcggcgc tgaaaccgtg aggcccggac cacaggctcc
       61 agatggaccc tgggaaggac aaagaggggg tgccccagcc ctcagggccg ccagcaagga
      121 agaaatttgt gataccctc gacgaggatg aggtccctcc tggagtggcc aagcccttat
      181 tccgatctac acagagcctt cccactgtgg acacctcggc ccaggcggcc cctcagacct
      241 acgccgaata tgccatctca cagcctctgg aaggggctgg ggccacgtgc cccacagggt
```

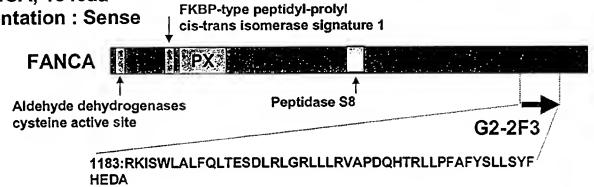
FIG. 18 (2/2)

```
 301 cagagcccct ggcaggagag acgcccaacc aggccctgaa acccggggca aaatccaaca
 361 gcatcattgt gagccctcgg cagaggggca atcccgtact gaagttcgtg cgcaatgtgc
 421 cctgggaatt tggcgacgta attcccgact atgtgctggg ccagagcacc tgtgccctgt
 481 tcctcagcct ccgctaccac aacctgcacc cagactacat ccatgggcgg ctgcagagcc
 541 tggggaagaa cttcgccttg cgggtcctgc ttgtccaggt ggatgtgaaa gatccccagc
 601 aggccctcaa ggagctggct aagatgtgta tcctggccga ctgcacattg atcctcgcct
 661 ggagccccga ggaagctggg cggtacctgg agacctacaa ggcctatgag cagaaaccag
 721 cggacctcct gatggagaag ctagagcagg acttcgtctc ccgggtgact gaatgtctga
 781 ccaccgtgaa gtcagtcaac aaaacggaca gtcagaccct cctgaccaca tttggatctc
 841 tggaacagct catcgccgca tcaagagaag atctggcctt atgcccaggc ctgggccctc
 901 agaaagcccg gaggctgttt gatgtcctgc acgagccctt cttgaaagta ccctgatgac
 961 cccagctgcc aaggaaaccc ccagtgtaat aataaatcgt cctcccaggc caggctcctg
1021 ctggc
//
```

Revised: October 24, 2001.

G2-2F3 / Fanconi Anemia Group A (FANCA)

The G2-2F3 sequence is identical to Fanconi Anemia Group A, FANCA, 1340aa
Orientation : Sense



Pfam HMM search was done at the Washington University web site

Aldehyde dehydrogenases cysteine active site (3-14): It is found in a nuclear protein associated with cell proliferation
FKBP-type peptidyl-prolyl cis-trans isomerase signature 1(159-175): One of two signature patterns for FKBP
PX(189-320): Novel domains in NADPH oxidase subunits, sorting nexins, and PI3-kinases: binding partners of SH3 domains?
PeptidaseS8(660-688): Subtilase family motif

FIG. 19

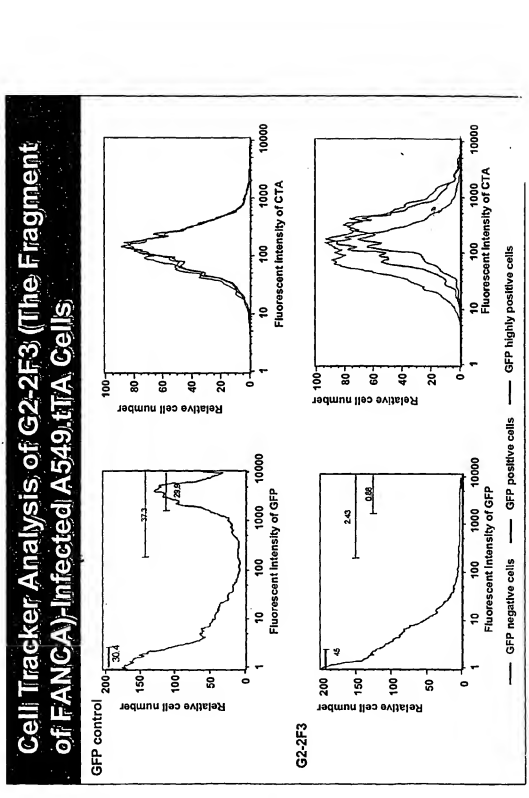

Cell Tracker Analysis of G2-2F3 (The Fragment of FANCA)-Infected A549-ETA Cells
GFP control
G2-2F3
Relative cell number
Fluorescent Intensity of GFP
Fluorescent Intensity of CTA
30.4
37.3
23.9
2.43
0.86
GFP negative cells
GFP positive cells
GFP highly positive cells

FIG. 20

G3-2H6 // DEAD/H Box Polypeptide 9 (DDX9)

The G3-2H6 sequence is identical to DEAD/H box polypeptide 9 (DDX9), 1279aa
Orientation: Antisense

C-teminus of GFP

GAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGAGGAGGCC
GCCAAGGCC
E F V T A A G I T L G M D E L Y K E E A A K A
GGTGGCAGCGGTGGCT*CCAGTGTGCTGGAAAG*CGCCACCTCCTCTTCCCTGTCCAAAGTA
GCCAGTTCC
G G S G G S S V L E S A T S S S L S K V A S S
ATAGGCCCCCCTACCACCTCCTCGCTGGAATCCCCAGATCCTCTGTAGCCTCCACTAGGC
CCTCTGTA
I G P P T T S S L E S P R S S V A S T R P S V
GTCTCCTCCAGAGTTGCCTCTAAAGCCACCTCGGGAGACTCCTCTATAGCCTCCACCAACA
CCTGCACC
V S S R V A S K A T S G D S S I A S T N T C T
ATATCCTGCCCGAAAGGAGTTGGCGCTGCCACCATAGCCTCCGCTACCATAGCCTCCACTG
CTATAGCC
I S C P K G V G A A T I A S A T I A S T A I A
ACCGCATAGCCTCCACCACTGTAACTAGAACCTCCCCTTCTATATCCGCTTCCATTGTCGTA
TCGGGCC
T A Stop (93 amino acids)
ATCTTGGCAGGACGTGGACCATCTCCATGCCGTGTACTGCCAATCATAA[illegible]
[illegible]
[illegible]TTCAGCATACGTTCATTTACGGGGTCCAACTGGCTGAT
GATAGCA

RB(4-69, 181-250): Double-stranded RNA binding motif
DEAD (486-546): DEAD/DEAH box helicase
HelC (670-679): Helicase conserved C-terminal domain
CLN3 (994-1007): CLN3 protein

CCAGTGTGCTGGAAAG,
CTTTCCAGCACACTGG
= BstXI linker

FIG. 21

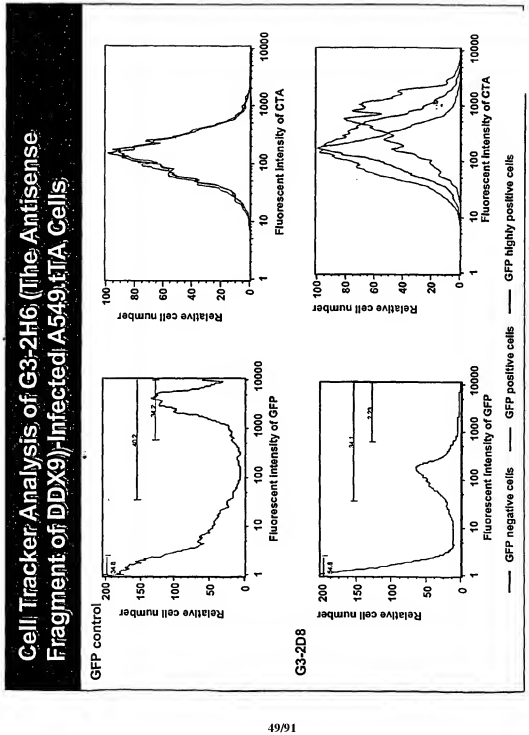
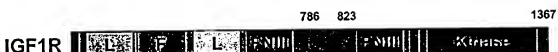

Cell Tracker Analysis of G3-2H6 (The Antisense Fragment of DDX9) Infected A549 LTA Cells
GFP control
G3-2D8
Relative cell number
Fluorescent Intensity of GFP
Fluorescent Intensity of CTA
GFP negative cells
GFP positive cells
GFP highly positive cells

FIG. 22

G3-2H2_1/ Insulin-like Growth Factor 1 Receptor (IGF1R)

The G3-2H2_1 sequence is identical to Insulin-like growth factor 1 receptor (IGF1R)

Orientation: Sense



786:ERTVISNLRPFTLYRIDIHSCNHEAEKLGCSASNFV
FA

Leader sequence (1-30)
L (51-172, 352-472): Receptor L domain, the L domains from insulin-like growth factor receptors make up the bilobal ligand binding site.
F (175-333): Furin-like cysteine rich region, which involves receptor aggregation
FNIII(489-587, 835-917): Fibronectin type III domain, the majority of which are involved in cell surface binding in some manner, or are receptor protein tyrosine kinases, or cytokine receptors.
Transmembrane (936-958)
Kinase(999-1266): Protein tyrosine kinase catalytic domain

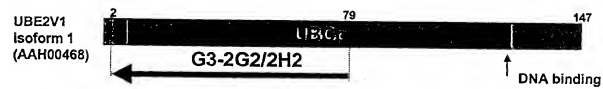
FIG. 23

Cell Tracker Analysis of G3-2H2_1 (The Fragment of IGF1R)-Infected A549-tTA Cells
GFP control
G3-2H2_1
Relative cell number
Fluorescent intensity of GFP
Fluorescent intensity of CTA
GFP negative cells
GFP positive cells
GFP highly positive cells

FIG. 24

G3-2G2 / 2H2 Ubiquitin-conjugating enzyme E2 variant 1 (UBE2V1)

The G3-2D8 sequence is identical to Ubiquitin-conjugating enzyme E2 variant 1 147aa Orientation: Antisense



233 bp insert

GGTGGCAGCGGTGGCT*CCAGTGTGCTGGAAAG*GTGCTTCTGGGTATTTAGGTCCACATT
CTATTTTAAG
G G S G G S S V L E R C F W V F R S T F Y F K
GCTGTATATTCGGTTTTCATAAATTGTTCTTGGAGGCCCAATTATCATCCCTGTCCATCTT
GTAAGTGT
A V Y S V F I N C S W R P N Y H P C P S C K C
CATGTCTTCGTCATCTTCTAGACCCCAGCTAACTGTGCCATCTCCTACTCCTTTCTGGCCT
TCTTCGAG
UBCc (15-141) = ubiquitin-conjugating enzyme E2, catalytic domain homolog [*Homo sapiens*]
TTCTTCCAAAAGTCGGAAATTCCGAGGGACTTTTACTCCCGAGCCCGTGGTGGCT*CTTTC*
CAGCACAGT
GCCAGTGTGCTGGAAAG,
CTTTCCAGCACAGTGG = BstXI linker

FIG. 25

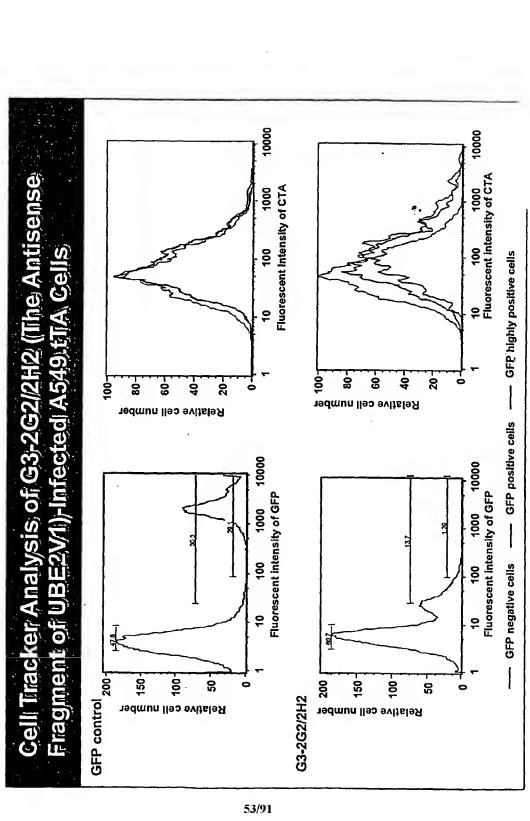

Cell Tracker Analysis of G3-2G2/2H2 (The Antisense Fragment of UBE2V1)-Infected A549 tTA Cells
GFP control
G3-2G2/2H2
Relative cell number
Fluorescent intensity of GFP
Fluorescent intensity of CTA
GFP negative cells
GFP positive cells
GFP highly positive cells

FIG. 26

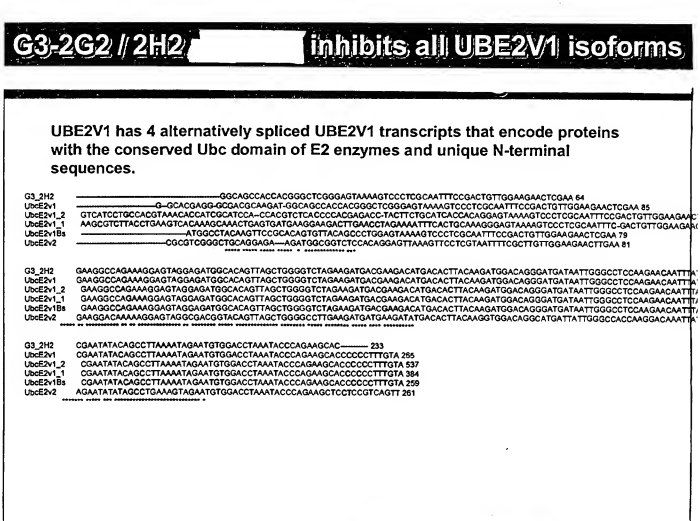

G3-2G2 / 2H2 inhibits all UBE2V1 isoforms
UBE2V1 has 4 alternatively spliced UBE2V1 transcripts that encode proteins with the conserved Ubc domain of E2 enzymes and unique N-terminal sequences.
G3_2H2 ——————————GGCAGCCACCACGGGCTCGGGAGTAAAAGTCCCTCGCAATTTCCGACTGTTGGAAGAACTCGAA 64
UbcE2v1 ——————————G—GCACGAGG-GCCACGCAAGAT-GGCAGCCACCACGGGCTCGGGAGTAAAAGTCCCTCGCAATTTCCGACTGTTGGAAGAACTCGAA 85
UbcE2v1_2 GTCATCCTGCCACGTAAACACCATCGCATCCA—CCACGTCTCACCCACGAGACCTACTTCTGCATCACCACAGGAGTAAAAGTCCCTCGCAATTTCCGACTGTTGGAAGAAC
UbcE2v1_1 AAGCGTCTTACCTGAAGTCACAAAGCAAACTGAGTGATGAAGGAAGACTTGAACCTAGAAAATTTCACTGCAAAGGGAGTAAAAGTCCCTCGCAATTTC-GACTGTTGGAAGAA
UbcE2v1Bs ——————————ATGGCCTACAAGTTCCGCACAGTGTTACAGCCCTGGAGTAAAAGTCCCTCGCAATTTCCGACTGTTGGAAGAACTCGAA 79
UbcE2v2 ——————————CGCGTGGGCTGCAGGAGA—AGATGGCGGTCTCCACAGGAGTTAAAGTTCCTCGTAATTTTCGCTTGTTGGAAGAACTTGAA 81
G3_2H2 GAAGGCCAGAAAGGAGTAGGAGATGGCACAGTTAGCTGGGGTCTAGAAGATGACGAAGACATGACACTTACAAGATGGACAGGGATGATAATTGGGCCTCCAAGAACAATTTA
UbcE2v1 GAAGGCCAGAAAGGAGTAGGAGATGGCACAGTTAGCTGGGGTCTAGAAGATGACGAAGACATGACACTTACAAGATGGACAGGGATGATAATTGGGCCTCCAAGAACAATTTA
UbcE2v1_2 GAAGGCCAGAAAGGAGTAGGAGATGGCACAGTTAGCTGGGGTCTAGAAGATGACGAAGACATGACACTTACAAGATGGACAGGGATGATAATTGGGCCTCCAAGAACAATTTA
UbcE2v1_1 GAAGGCCAGAAAGGAGTAGGAGATGGCACAGTTAGCTGGGGTCTAGAAGATGACGAAGACATGACACTTACAAGATGGACAGGGATGATAATTGGGCCTCCAAGAACAATTTA
UbcE2v1Bs GAAGGCCAGAAAGGAGTAGGAGATGGCACAGTTAGCTGGGGTCTAGAAGATGACGAAGACATGACACTTACAAGATGGACAGGGATGATAATTGGGCCTCCAAGAACAATTTA
UbcE2v2 GAAGGACAAAAAGGAGTAGGCGACGGTACAGTTAGCTGGGGCCTTGAAGATGATGAAGATATGACACTTACAAGGTGGACAGGCATGATTATTGGGCCACCAAGGACAAATTTA
G3_2H2 CGAATATACAGCCTTAAAATAGAATGTGGACCTAAATACCCAGAAGCAC——— 233
UbcE2v1 CGAATATACAGCCTTAAAATAGAATGTGGACCTAAATACCAGAAGCACCCCCCTTTGTA 255
UbcE2v1_2 CGAATATACAGCCTTAAAATAGAATGTGGACCTAAATACCCAGAAGCACCCCCCTTTGTA 537
UbcE2v1_1 CGAATATACAGCCTTAAAATAGAATGTGGACCTAAATACCCAGAAGCACCCCCCTTTGTA 384
UbcE2v1Bs CGAATATACAGCCTTAAAATAGAATGTGGACCTAAATACCCAGAAGCACCCCCTTTGTA 259
UbcE2v2 AGAATATATAGCCTGAAAGTAGAATGTGGACCTAAATACCCAGAAGCTCCTCCGTCAGTT 261

FIG. 27

FIG. 28 (1/2)

SEQ ID NO:29
Size: 181
DNA FANCA

CCAGTGTGCTGGAAAGGAGGAAGATATCCTGGCTGGCACTCTTTCAGTTGACAGAGAGTGACCTCAGGCTGGGGC
GGCTCCTCCTCCGTGTGGCCCCGGATCAGCACACCAGGCTGCTGCCTTTCGCTTTTTACAGTCTTCTCTCCTACT
TCCATGAAGACGCGGCTTTCCAGCACAGTGG

SEQ ID NO:30
Size: 603
DNA DDX9

CCAGTGTGCTGGAAAGCGCCACCTCCTCTTCCCTGTCCAAAGTAGCCAGTTCCATAGGCCCCCCTACCACCWCCT
CGCTGGAATCCCCCAGATCCTCTGTAGCCTCCACTAGGCCCTCTGTAGTCTCCTCCAGAGTTGCCTCTAAAGCCA
CCTCGGGAGACTCCTCTATAGCCTCCACCAACACCTGCACCATATCCTGCCCGAAAGGAGTTGGCGCTGCCACCA
TAGCCTCCGCTACCATAGCCTCCACTGCTATAGCCACCGCATAGCCTCCACCACTGTAACTAGAACCTCCCCTTC
TATATCCGCTTCCATTGTCGTATCGGGCCATCTTGGGAGGACGTGGACCATCTCCATGCCGTGTACTGCCAATCA
TAAGGTTGATACCAGCAGCTGAGGGTCTAGAGATCTGACGGATCATGTTCAGCATACGTTCATTTACGGGGTCCA
ACTGGCTGATGATAGCAGGTTGTTTGGTTACTTCAACAACCAAAGCCTCCATGGCTGCCCGGAGACCAGTGATAC
AGGCAGCAGCTTCATGAGATATTTGCAGTTTAATCCAGTCATCTACAAGCACAATCTGCCCACTTTCCAGCACAG
TGG

SEQ ID NO:31
Size: 145
DNA IGF1R

CCAGTGTGTTGGAAAGGGAGAGAACTGTCATTTCTAACCTTCGGCCTTTCACATTGTACCGCATCGATATCCACA
GCTGCAACCACGAGGCTGAGAAGCTGGGCTGCAGCGCCTCCAACTTCGTCTTTGCTTTCCAGCACAGTGG

SEQ ID NO:32
Size: 269
DNA UBEV2V1

CCAGTGTGCTGGAAAGGTGCTTCTGGGTATTTAGGTCCACATTCTATTTTAAGGCTGTATATTCGGTTTTCATAA
ATTGTTCTTGGAGGCCCAATTATCATCCCTGTCCATTCTTGTAAGATGTCATGTCTTCGTCATCTTCTAGACCCCA
GCTAACTGTGCCATCTCCTACTCCTTTCTGGCCTTCTTCGAGATTCCTCCAACAGTCGGAAATTGCGAGGGACTT
TATACATCCCGAGCCCGTGGTGGCTGCCCTTTCCAGCACACTGG

SEQ ID NO:33
Size: 499
DNA aldehyde dehydrogenase

CCAGTGTGCTGGAAAGGAGCAAACTCCTCTCACTGCTCTCCACGTGGCATCTTTAATAAAAGAGGCAGGGTTTCC
TCCTGGAGTAGTGAATATTGTTCCTGGTTATGGGCCTACAGCAGGGGCAGCCATTTCTTCTCACATGGATATAGA
CAAAGTAGCCTTCACAGGATCAACAGAGGTTGGCAAGTTGATCAAAGAAGCTGCCGGGAAAAGCAATCTGAAGAG
GGTGACCCTGGAGCTTGGAGGAAAGAGCCCTTGCATTGTGTTAGCTGATGCCGACTTGGACAATGCTGTTGAATT
TGCACACCATGGGGTATTCTACCACCAGGGCCAGTGTTGTATAGCCGCATCCAGGATTTTTGTGGAAGAATCAAT
TTATGATGAGTTTTGTTCGAAGGAGTGTTGAGCGGGCTAAGAACGTATATCCTTGGAAACATCCTCTGACCCCAG
GAGTCACTCAAAGGCCCTCAGATTGACAAGGACTTTCCAGACACAGTGG

SEQ ID NO:34
Size: 425
DNA pyruvate kinase

FIG. 28 (2/2)

CCAGTGTGCTGGAAAGGCTGCCCACTTCCACCACCTTGCAGATGTTCTTGTAGTCCAGCCACAGGATGTTCTCGT
CACACTTTTCCATGTAGGCGTTATCCAGCGTGATTTTGAGAGTGGCTCCCTTCTTCAGCTCCACCTCTGCAGTGC
CGCTGCCCTTGATGAGCCCAGTTCGGATCTCAGGTCCTTTAGTGTCTAGAGCCACAGCAACGGGCCGGTAGAGGA
TGGGGTCAGAAGCAAAGCTTTCCGTGGCTGTGCGCACATTCTTGATGGTCTCCGCATGGTACTCATGAGTTCCAT
GAGAGAAGTTCAGACGAGCCACATTCATTCCAGACTTAATCATCTCCTTCAACGTCTCCACTGGATCGGGAAGCT
GGGCCAATGGTACAGATGATGCCAGTGTTCCGGGCTTTCCAGCACAGTGG

SEQ ID NO:35
Size:
DNA G6PD

CCAGTGTGCTGGAAACTTTCCAGTTCTCCATGGCCACCANACACAGCATCTGCAGTAGGTGGTTCTGCATCACGT
CCCGGATGATCCCAAATTCATCGAAATAGCCCCCGCGACCCTCAGTGCCAAAGGGCTCCTTGAAGGTGAGGATAA
CGCAGGCGATGTTGTCCCGGTTCCANATGGGGCCGAAGATCCTGTTGGCAAATCTCAGCACCATGAGGTTCTCTT
TCCAGCACAGTGG

Dominant Negative Mutants of BAP-1

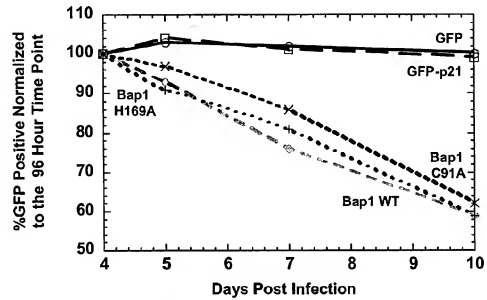
Point mutants: C91A, H169A- catalytic residues in the protease domain.
(EMBO J. 1997 Jul 1;16(13):3787-96. PMID: 9233788)

```
CLUSTAL W (1.8) multiple sequence alignment

Uch-l3          MEGQRWLPLEANPEVTNQFLKQLGLHPNWQFVDVYGMDPELLSMVPRPVCAVLLLFPITE
BAP-1            MNKGWLELESDPGLFTLLVEDFGVKG-VQVEEIY----DLQSKCQGFVYGFIFLFKWIE
                  .: ** **::* : . :::::*::    *. ::*    :* *    ** ..::**   *

                                                  ↓
Uch-l3          KYEVFR--TEEEEKIKSQGQDVTSSVYFMKQTISNACGTIGLIHAIANNKDKMHFESGST
BAP-1           ERRSRRKVSTLVDDTSVIDDDIVNNMFFAHQLIPNSCATHALLSVLLNCSS---VDLGPT
                : .  *  :   :. .  .:*:...::* :* *.*:*.* .*: .: * ..   .: *.*

                                                                  ↓
Uch-l3          LKKFLEESVSMSPEERARYLENYDAIRVTHETSAHEGQTEAP-------SIDEKVDLHFI
BAP-1           LSRMKDFTKGFSPESKGYAIGNAPELAKAHNSHARPEPRHLPEKQNGLSAVRTMEAFHFV
                *.:: : : .:***.:.  : *   :  :*:: *:    . *        ::      :**:

Uch-l3          ALVHVDGHLYELDGRKPFPINHGETS-DETLLEDAIEVCKKFMERDPD-----ELRFNAI
BAP-1           SYVPITGRLFELDGLKVYPIDHGPWGEDEEWTDKARRVIMERIGLATAGEPYHDIRFNLM
                : * : *:*:**** * :**:**  . **   :.* .*  : :   .       ::*** :
```

Bold: Catalytic residue

FIG. 29

Expression of Bap1 WT and Protease Mutants is Antiproliferative in HeLa Cells



%GFP Positive at 96 Hours

GFP	80.8
GFP-p21	19.5
Bap1 WT	12.0
Bap1 C91A	14.7
Bap1 H169A	12.5

RIGEL

FIG. 30

Expression of Bap1 WT Protein is Antiproliferative in HeLa Cells in the Celltracker Assay

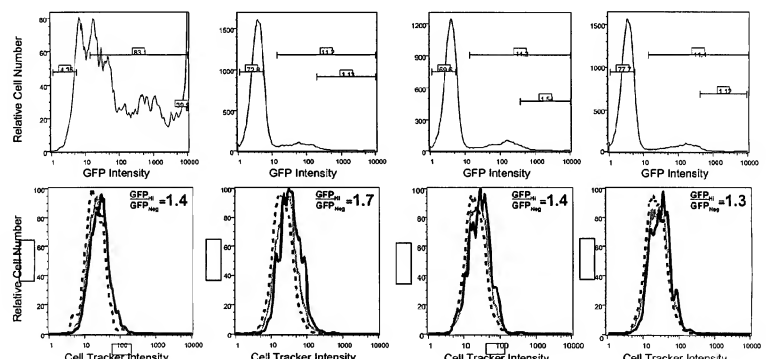
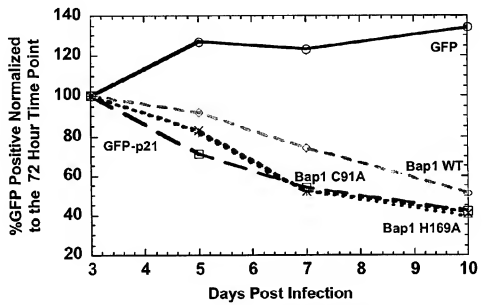


FIG. 31

Expression of Bap1 Protease Mutants is Slightly More Antiproliferative Than Expression of Bap1 WT in H1299 Cells



%GFP Positive at 72 Hours

GFP	49.1
GFP-p21	4.2
Bap1 WT	5.0
Bap1 C91A	6.8
Bap1 H169A	7.1

FIG. 32

Expression of Bap1 WT and Bap1 Protease Mutants is Antiproliferative in H1299 Cells in the Celltracker Assay

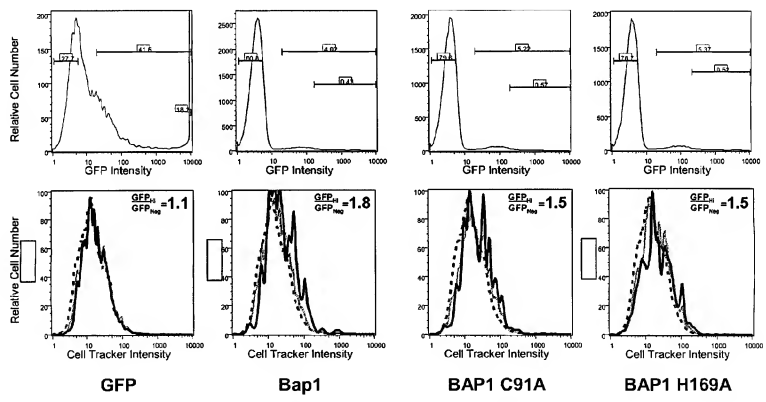


FIG. 33

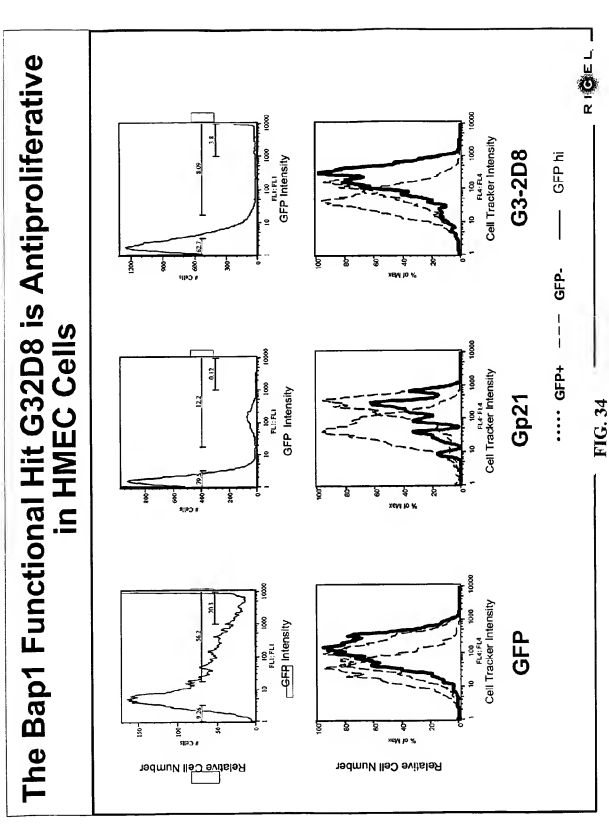

The Bap1 Functional Hit G32D8 is Antiproliferative in HMEC Cells
Relative Cell Number
GFP Intensity
Cell Tracker Intensity
GFP
Gp21
G3-2D8
..... GFP+
---- GFP-
—— GFP hi

FIG. 34

The Bap1 Functional Hit G3-2D8 is Antiproliferative in PrEC Cells

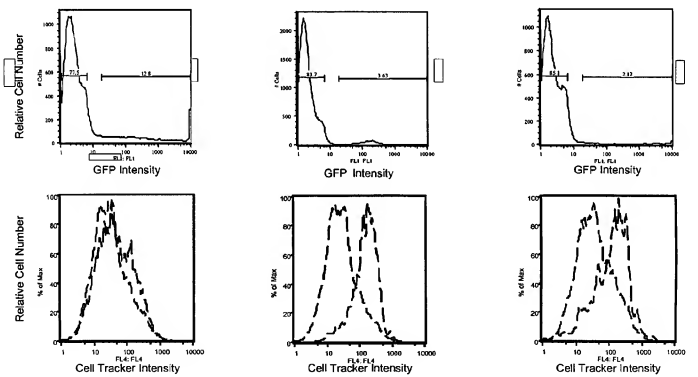
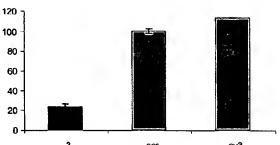


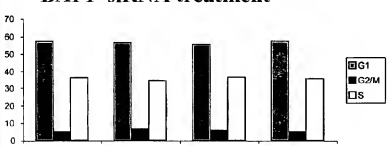
FIG. 35

BAP1 Specific siRNA Has an Antiproliferative Effect on HeLa Cells

BAP1 mRNA levels in HeLa after siRNA treatment (Taqman)



HeLa cell cycle profile after BAP1 siRNA treatment

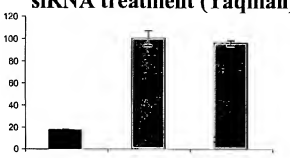


BrdU incorporation by HeLa treated with BAP1 siRNA

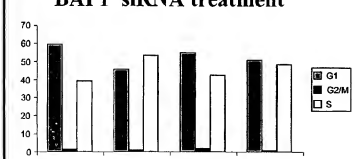
FIG. 36

BAP1 Specific siRNA Induces G1 Arrest in H1299 Cells

BAP1 mRNA levels in H1299 after siRNA treatment (Taqman)



H1299 cell cycle profile after BAP1 siRNA treatment



BrdU incorporation by H1299 treated with BAP1 siRNA

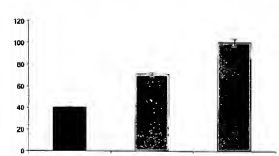


FIG. 37

RIGEL

Soluble GST-Bap1 Protein can be Expressed from SF9 Cells

GST-Bap1 was produced using the baculovirus transfer vector pDEST20 along with the Bac-to-Bac baculovirus expression system (invitrogen). GST-Bap1(1) and GST-Bap1(2) refer to two different virus dilutions used for expression.

FIG. 38

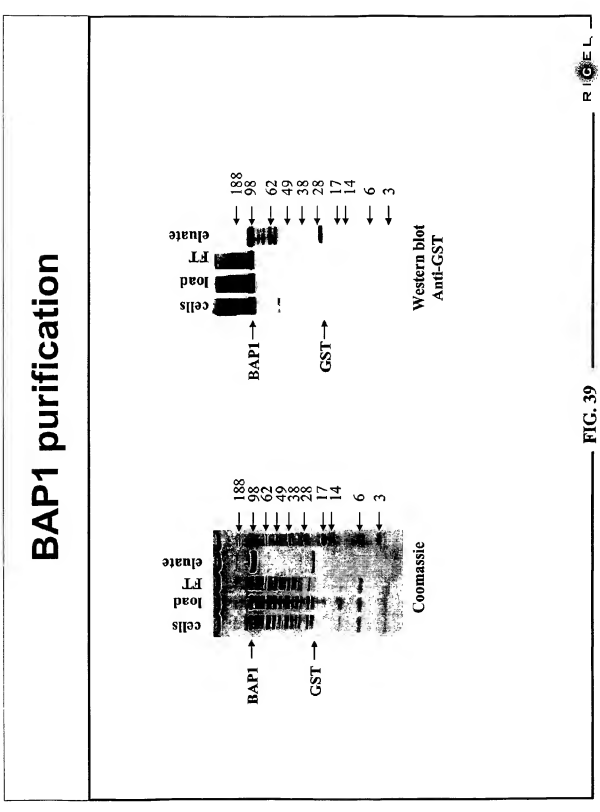
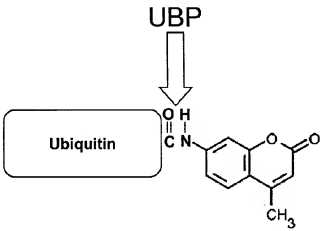

BAP1 purification
cells
load
FT
elute
BAP1
GST
Coomassic
188
98
62
49
38
28
17
14
6
3
Western blot
Anti-GST
R I G E L

FIG. 39

Example of Fluorogenic Ub Cleavage Assay

Aminomethyl-coumarin cleavage from Ub C-terminus generates fluorescence emission in solution-phase assay



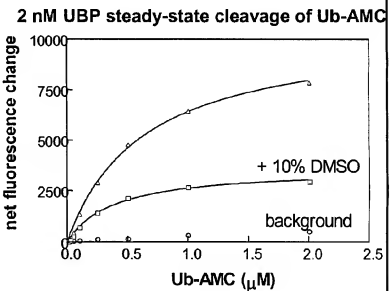
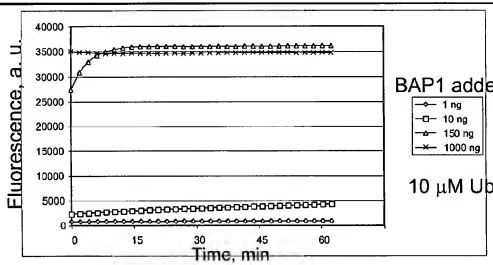


FIG. 40

BAP1 is an Active Ubiquitin Protease



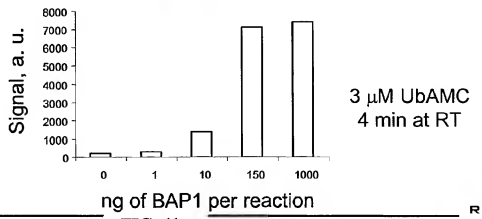


FIG. 41

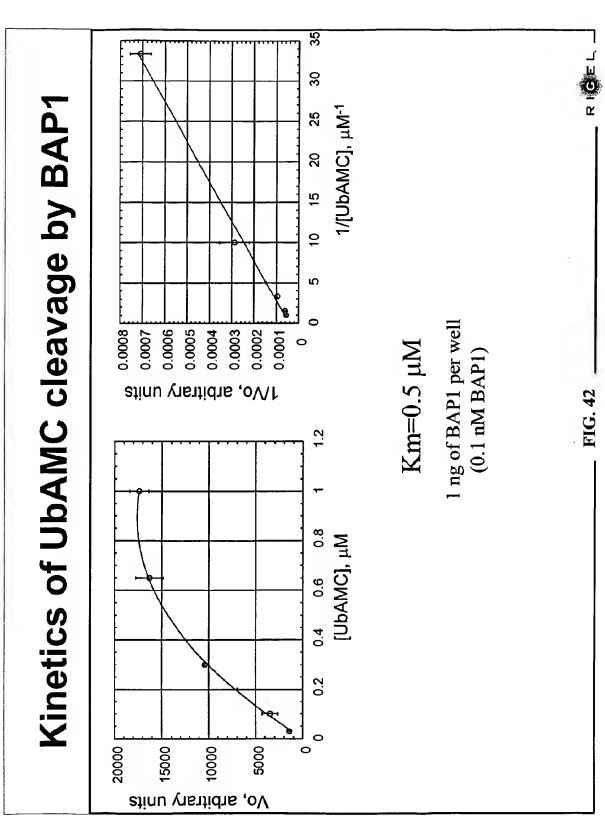
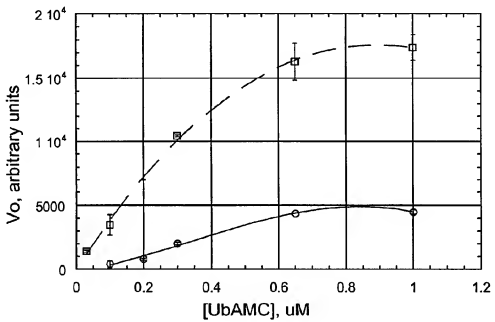

Kinetics of UbAMC cleavage by BAP1
Vo, arbitrary units
20000
15000
10000
5000
0
0
0.2
0.4
0.6
0.8
1
1.2
[UbAMC], µM
1/Vo, arbitrary units
0.0008
0.0007
0.0006
0.0005
0.0004
0.0003
0.0002
0.0001
0
0
5
10
15
20
25
30
35
1/[UbAMC], µM-1
Km=0.5 µM
1 ng of BAP1 per well
(0.1 nM BAP1)

FIG. 42

UbCHO Acts as Specific Inhibitor of BAP1



Ki=9 nM

1 ng of BAP1 per well
(0.1 nM BAP1)

FIG. 43

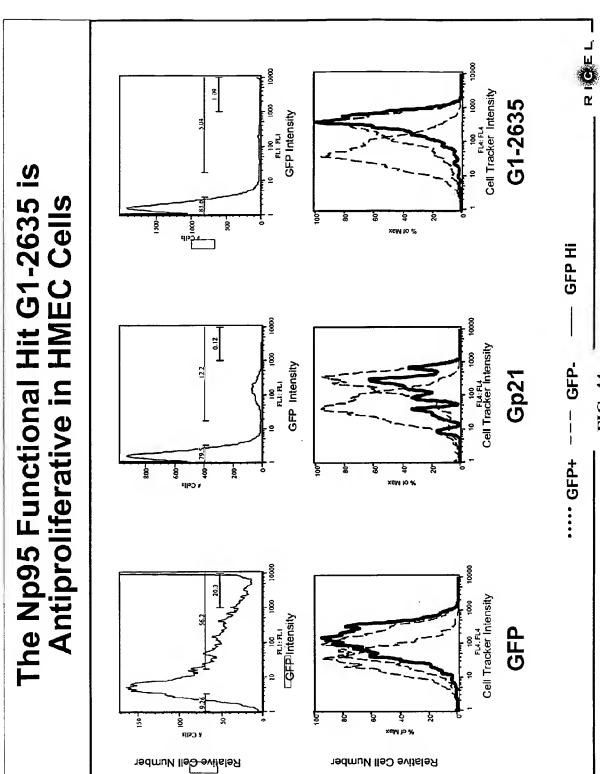

The Np95 Functional Hit G1-2635 is Antiproliferative in HMEC Cells
Relative Cell Number
GFP Intensity
Cell Tracker Intensity
GFP
Gp21
G1-2635
GFP+
GFP-
GFP Hi

FIG. 44

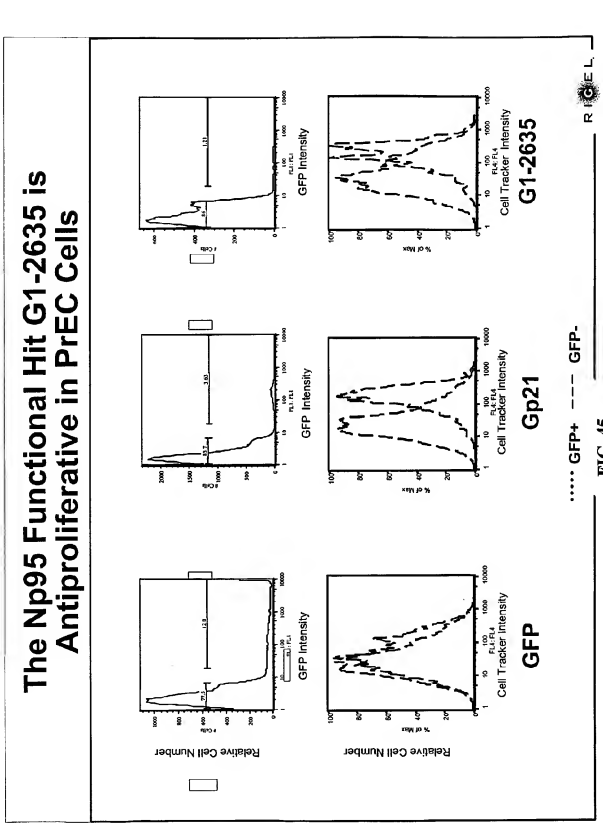

The Np95 Functional Hit G1-2635 is Antiproliferative in PrEC Cells
Relative Cell Number
GFP Intensity
Cell Tracker Intensity
% of Max
GFP
Gp21
G1-2635
..... GFP+ ---- GFP-

FIG. 45

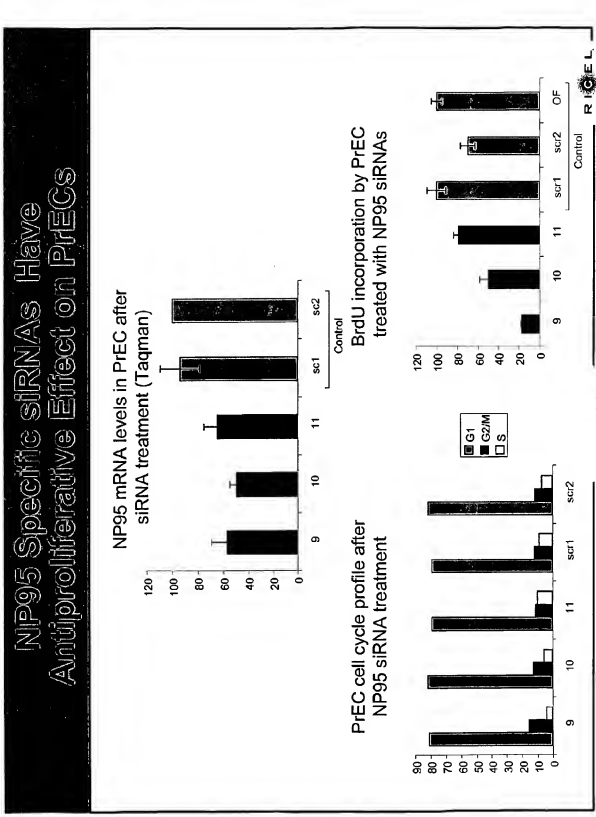

NP95 Specific siRNAs Have Antiproliferative Effect on PrECs
NP95 mRNA levels in PrEC after siRNA treatment (Taqman)
120
100
80
60
40
20
0
9
10
11
sc1
sc2
Control
PrEC cell cycle profile after NP95 siRNA treatment
90
80
70
60
50
40
30
20
10
0
G1
G2/M
S
scr1
scr2
BrdU incorporation by PrEC treated with NP95 siRNAs
OF
Control
RIGEL

FIG. 46

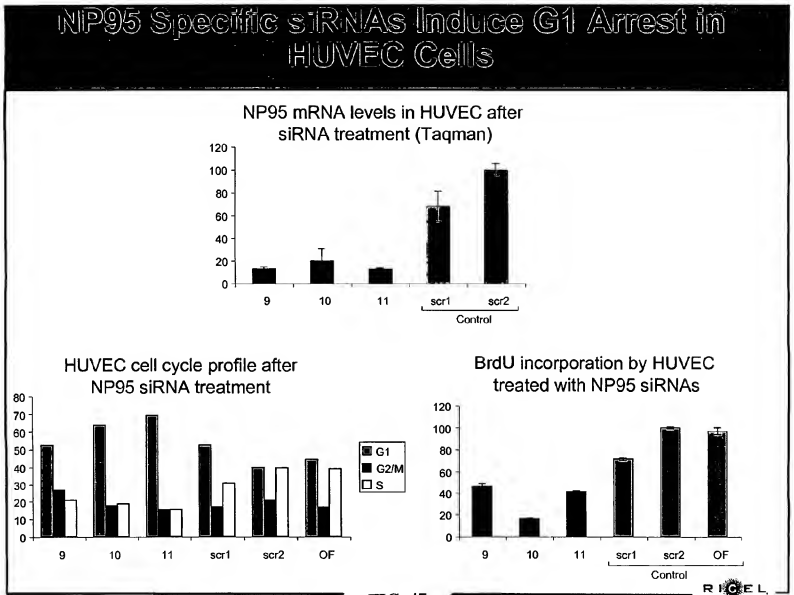
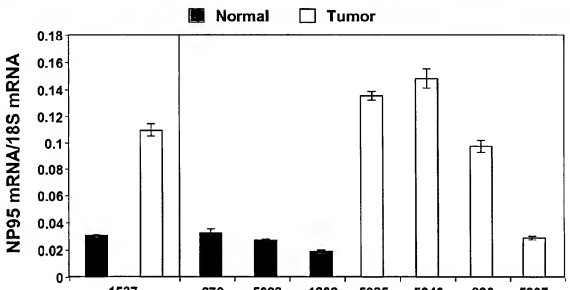

NP95 Specific siRNAs Induce G1 Arrest in HUVEC Cells
NP95 mRNA levels in HUVEC after siRNA treatment (Taqman)
120
100
80
60
40
20
0
9
10
11
scr1
scr2
Control
HUVEC cell cycle profile after NP95 siRNA treatment
80
70
60
50
40
30
20
10
0
9
10
11
scr1
scr2
OF
G1
G2/M
S
BrdU incorporation by HUVEC treated with NP95 siRNAs
120
100
80
60
40
20
0
9
10
11
scr1
scr2
OF
Control
RIGEL

FIG. 47

Taqman Analysis of NP95 mRNA Expression in Samples Obtained from Patients with Breast Carcinoma



Ductal Adenocarcinoma	836,1537, 5035, 5346
Lobular Adenocarcinoma	5297

N = 3, 20 ng total RNA/sample

RIGEL

FIG. 48

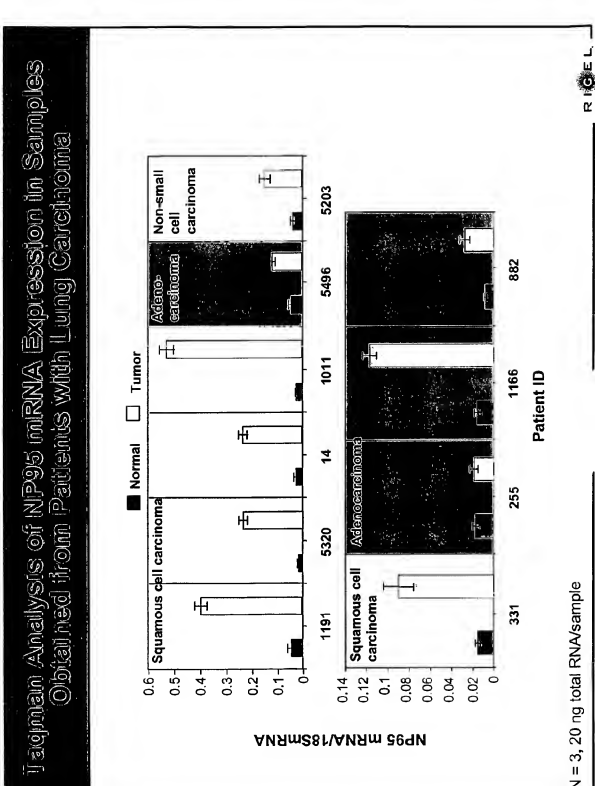
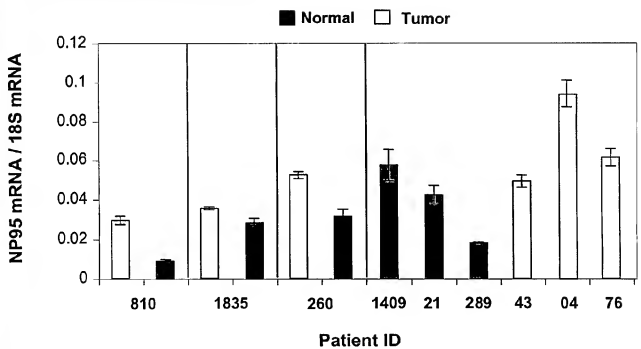

Taqman Analysis of NP95 mRNA Expression in Samples Obtained from Patients with Lung Carcinoma
Normal
Tumor
Squamous cell carcinoma
Adeno-carcinoma
Non-small cell carcinoma
NP95 mRNA/18SmRNA
0.6
0.5
0.4
0.3
0.2
0.1
0
1191
5320
14
1011
5496
5203
Squamous cell carcinoma
Adenocarcinoma
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
331
255
1166
882
Patient ID
N = 3, 20 ng total RNA/sample
ROCHE

FIG. 49

Taqman Analysis of NP95 mRNA Expression in Samples Obtained from Patients with Prostate Adenocarcinoma



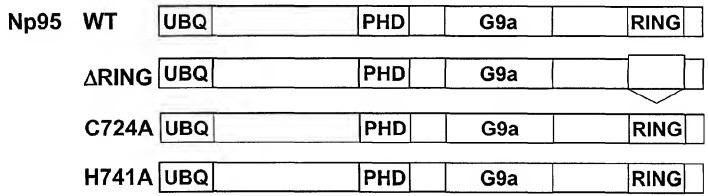
All tumors of acinar cell origin

N=3, 20 ng total RNA Sample

RIGEL

FIG. 50

Dominant Negative Mutants of Np95



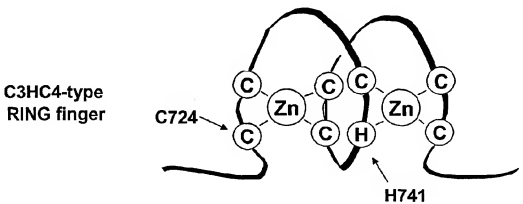
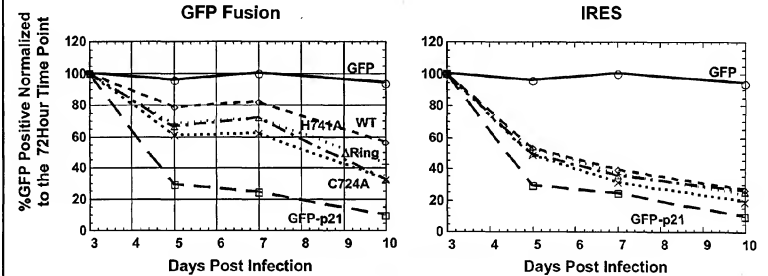


FIG. 51

GFP-fused Np95 Ring Finger Mutants are Slightly more Antiproliferative than GFP-fused Np95 WT in HCT116 Cells

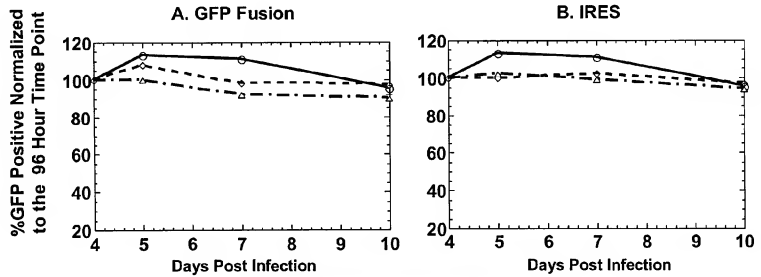


%GFP Positive at 72 Hours

	GFP Fusion	IRES
GFP	89.6	89.6
GFP-p21	38.8	38.8
Np95 WT	54.6	16.1
Np95 C724A	39.0	11.8
Np95 H741A	49.5	17.3
Np95 Δ Ring	41.3	19.0

FIG. 52

No Antiproliferative Effects Are Observed for Np95 WT and Ring Finger Mutant Constructs in A549 Cells



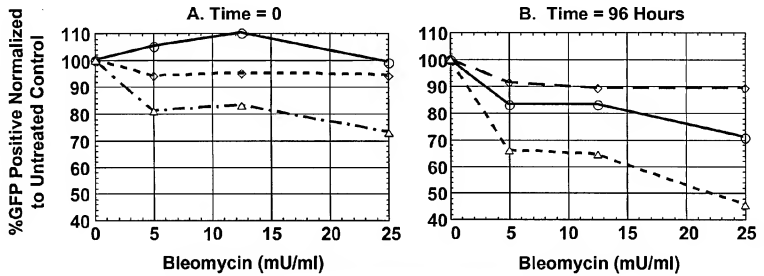
%GFP Positive at 96 Hours

	GFP Fusion	IRES
GFP	40.1	40.1
Np95 WT	8.6	23.9
Np95 Δ Ring	22.7	24.7

RIGEL

FIG. 53

A549 Cells Expressing GFP-Np95 ΔRing Become Sensitized to Bleomycin Treatment



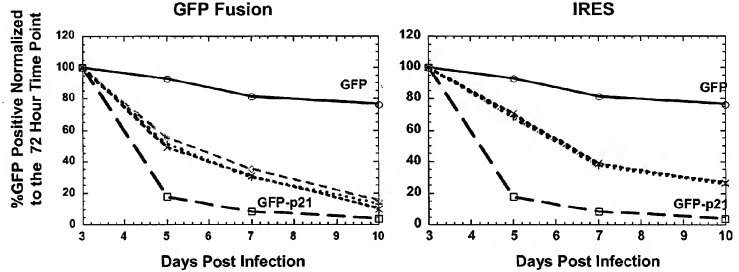
%GFP Positive at Time = 0

GFP	66.7
GFP-Np95 WT	20.5
GFP-Np95 Δ Ring	18.8

FIG. 54

RIGEL

Np95 WT and RING Finger Mutant Constructs are Strongly Antiproliferative in HMECs

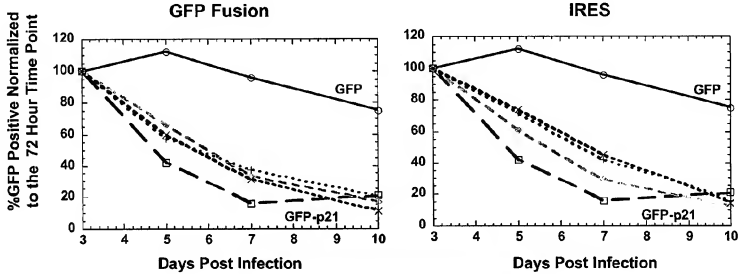


%GFP Positive at 72 Hours

	GFP Fusion	IRES
GFP	34.7	34.7
GFP-p21	7.1	7.1
Np95 WT	5.7	10.7
Np95 C724A	3.5	9.8
Np95 H741A	8.2	11.1

FIG. 55

Np95 WT and RING Finger Mutant Constructs are Strongly Antiproliferative in PrECs



%GFP Positive at 72 Hours

	GFP Fusion	IRES
GFP	10.6	10.6
GFP-p21	1.6	1.6
Np95 WT	1.7	2.7
Np95 C724A	0.91	2.3
Np95 H741A	2.2	2.4

FIG. 56

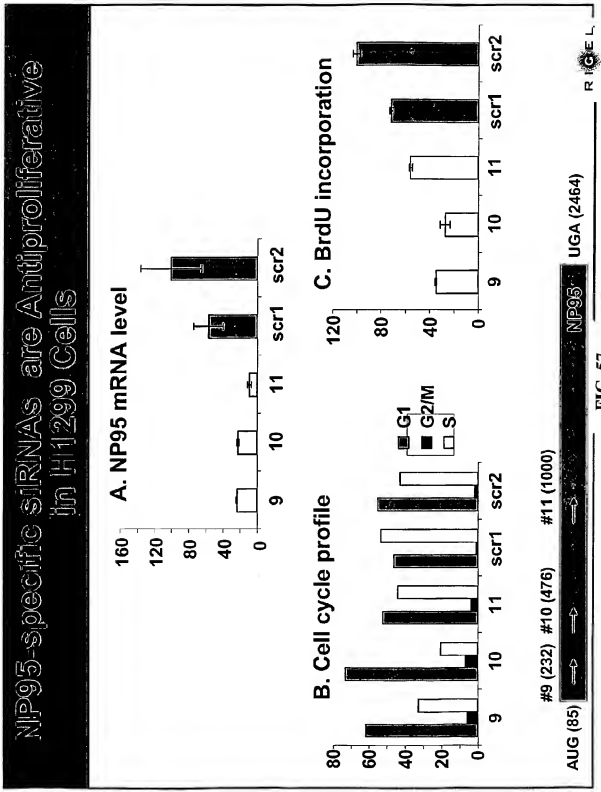

NP95-specific siRNAs are Antiproliferative in H1299 Cells
A. NP95 mRNA level
160
120
80
40
0
9
10
11
scr1
scr2
B. Cell cycle profile
80
60
40
20
0
G1
G2/M
S
C. BrdU incorporation
120
80
40
0
#9 (232)
#10 (476)
#11 (1000)
AUG (85)
NP95
UGA (2464)

FIG. 57

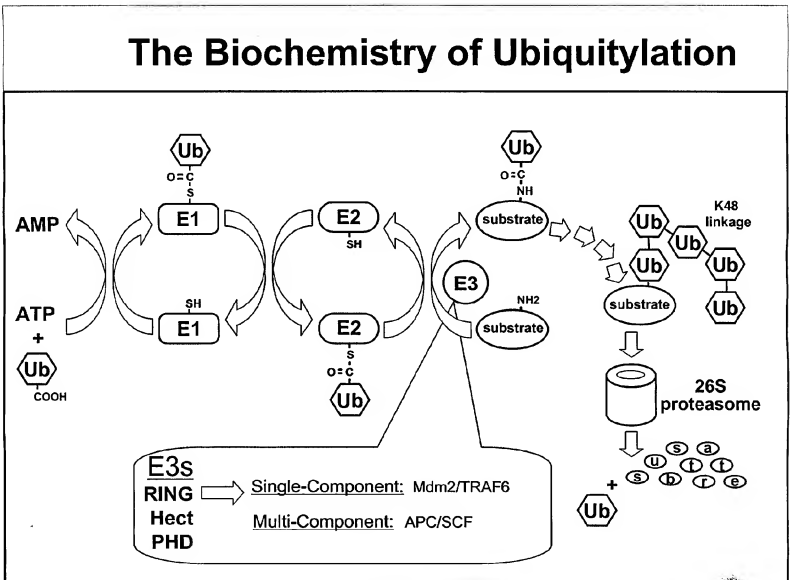

The Biochemistry of Ubiquitylation
Ub
O=C
S
AMP
E1
ATP
+
Ub
COOH
SH
E1
E2
SH
E2
S
O=C
Ub
E3
NH2
substrate
Ub
O=C
NH
substrate
K48
linkage
Ub
Ub
Ub
Ub
substrate
26S
proteasome
ubiquitin
+
Ub
E3s
RING
Hect
PHD
Single-Component: Mdm2/TRAF6
Multi-Component: APC/SCF

FIG. 58

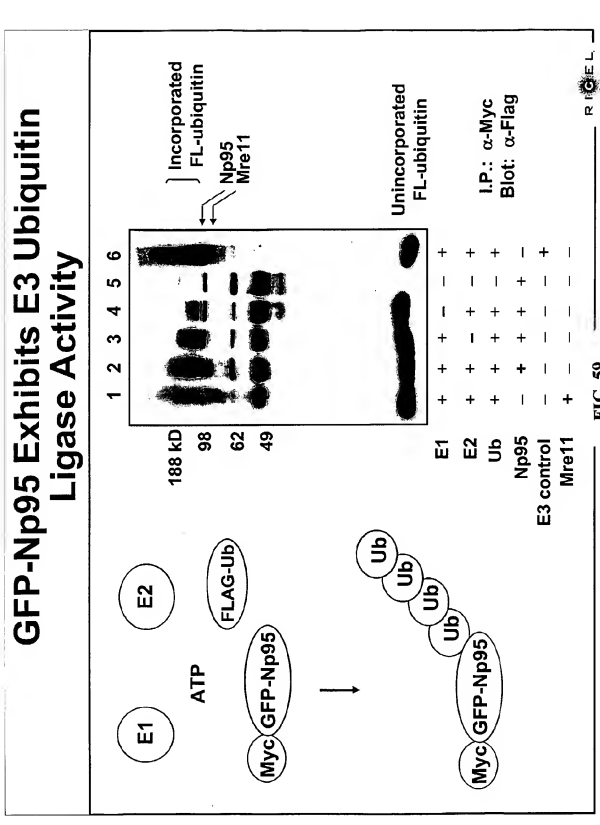
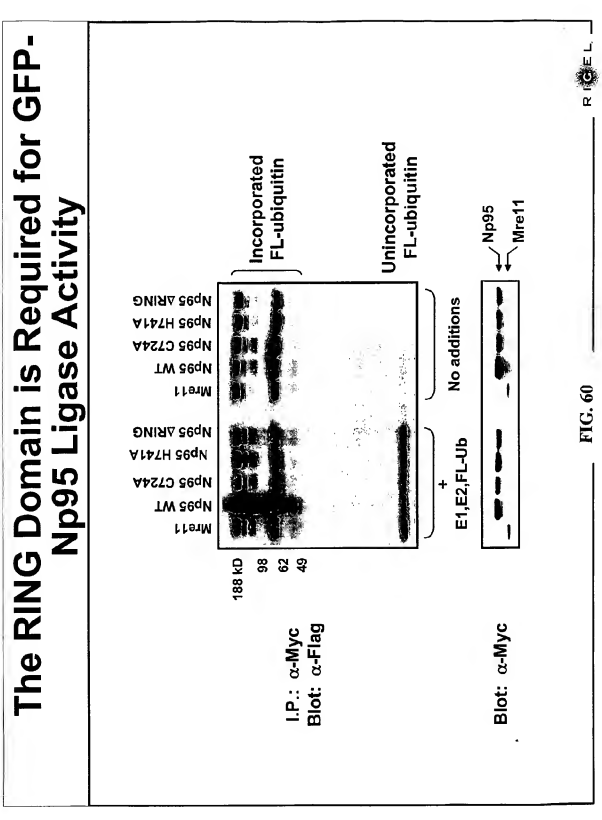

GFP-Np95 Exhibits E3 Ubiquitin Ligase Activity
E1
E2
ATP
FLAG-Ub
Myc GFP-Np95
Ub
Ub
Ub
Ub
Myc GFP-Np95
1 2 3 4 5 6
188 kD
98
62
49
Incorporated FL-ubiquitin
Np95
Mre11
Unincorporated FL-ubiquitin
E1 + + + – – +
E2 + + – + – +
Ub + + + + – +
Np95 – + + + + –
E3 control – – – – – +
Mre11 + – – – – –
I.P.: α-Myc
Blot: α-Flag

FIG. 59


The RING Domain is Required for GFP-Np95 Ligase Activity
I.P.: α-Myc
Blot: α-Flag
188 kD
98
62
49
Mre11
Np95 WT
Np95 C724A
Np95 H741A
Np95 ΔRING
Mre11
Np95 WT
Np95 C724A
Np95 H741A
Np95 ΔRING
Incorporated FL-ubiquitin
Unincorporated FL-ubiquitin
+
E1,E2,FL-Ub
No additions
Blot: α-Myc
Np95
Mre11
FIG. 60

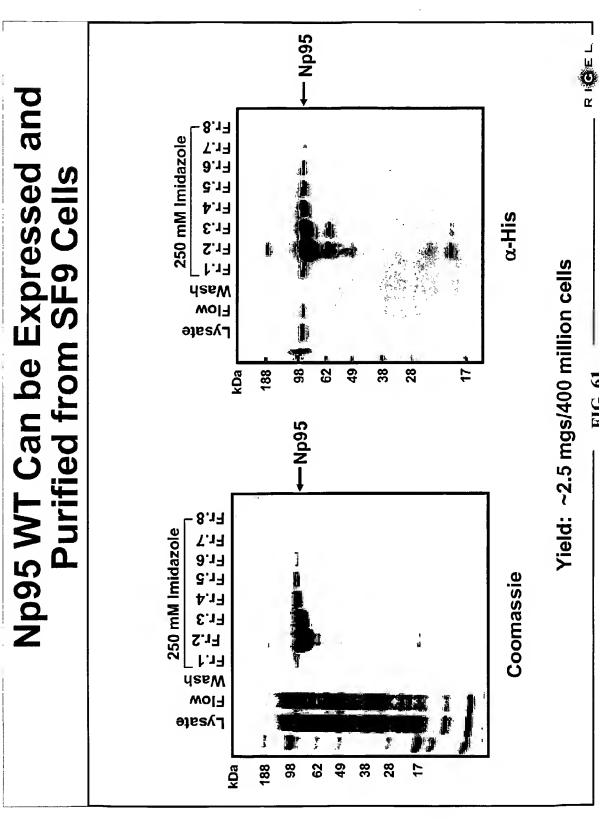

Np95 WT Can be Expressed and Purified from SF9 Cells
Lysate
Flow
Wash
250 mM Imidazole
Fr.1
Fr.2
Fr.3
Fr.4
Fr.5
Fr.6
Fr.7
Fr.8
kDa
188
98
62
49
38
28
17
Np95
Coomassie
α-His
Yield: ~2.5 mgs/400 million cells

FIG. 61

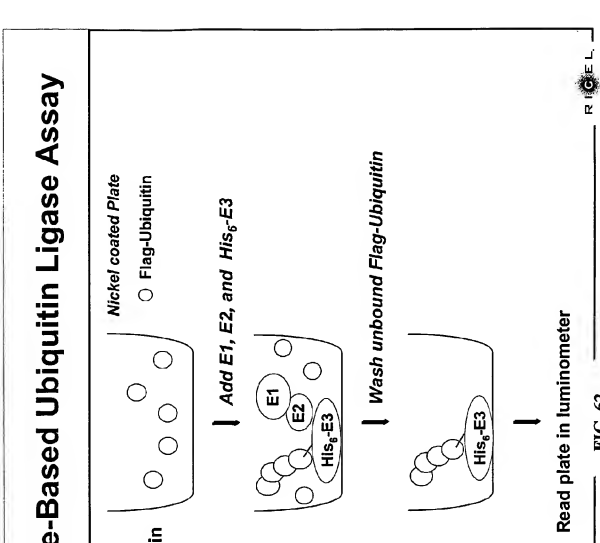
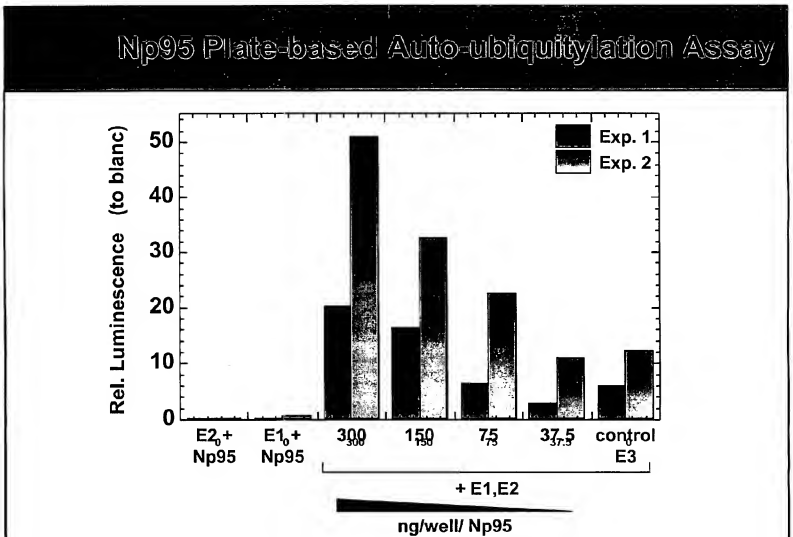

Rigel Plate-Based Ubiquitin Ligase Assay
Ligase assay reaction buffer containing Flag-Ubiquitin
Nickel coated Plate
Flag-Ubiquitin
Add E1, E2, and His6-E3
1 Hr. at room temp.
E1
E2
His6-E3
Wash unbound Flag-Ubiquitin
Add α-Flag HRP conjugate to detect the Flag-Ubiquitin
His6-E3
Read plate in luminometer
RIGEL

FIG. 62



Reactions contained 100 ng Fl-Ub, 5 ng of E1 and, 20 ng of E2 per well. The Np95 controls contained 150 ng Np95. The E3 control contained 75 ng E3. The two data sets are results of duplicate assays.

RIGEL

FIG. 63